. 08:37:38

STATE OF MINNESOTA DISTRICT COURT
COUNTY OF RAMSEY SECOND JUDICIAL DISTRICT

- -

THE STATE OF MINNESOTA,
BY HUBERT H. HUMPHREY, III,
ITS ATTORNEY GENERAL,

AND

BLUE CROSS AND BLUE SHIELD
OF MINNESOTA,

PLAINTIFFS,

VS. FILE NO. C1-94-8565

PHILIP MORRIS INCORPORATED, R.J.
REYNOLDS TOBACCO COMPANY, BROWN &
WILLIAMSON TOBACCO CORPORATION,
B.A.T. INDUSTRIES P.L.C., LORILLARD
TOBACCO COMPANY, THE AMERICAN
TOBACCO COMPANY, LIGGETT GROUP, INC.,
THE COUNCIL FOR TOBACCO RESEARCH-U.S.A.,
INC., AND THE TOBACCO INSTITUTE, INC.,

DEFENDANTS.

- -

DEPOSITION OF
SCOTT L. ZEGER, Ph.D.
VOLUME I

September 10, 1997
8:38 a.m.

REPORTED BY: JENNIFER S. SATI
REGISTERED PROFESSIONAL REPORTER
CERTIFIED REALTIME REPORTER
RAY J. LERSCHEN & ASSOCIATES
620 PLYMOUTH BUILDING
MINNEAPOLIS, MINNESOTA 55402

DEPOSITION OF SCOTT L. ZEGER, Ph.D., taken at the Law Offices of Robins, Kaplan, Miller & Ciresi, 2800 LaSalle Plaza, 800 LaSalle Street, Minneapolis, Minnesota, on the 10th day of September, 1997, commencing at 8:30 a.m., before Jennifer S. Sati, Notary Public.

* * * *

A P P E A R A N C E S

On Behalf of the Plaintiffs:

Robins, Kaplan, Miller & Ciresi
2800 LaSalle Plaza
800 LaSalle Avenue
Minneapolis, Minnesota 55402

BY: Thomas L. Hamlin

On Behalf of Philip Morris Incorporated:

Dorsey & Whitney
Pillsbury Center South
220 South Sixth Street
Minneapolis, Minnesota 55402-1498

BY: Robert A. Schwartzbauer

Arnold & Porter
555 Twelfth Street, N.W.
Washington, D.C. 20004-1202

BY: Thomas E. Silfen
Ellen Steury

On Behalf of R.J. Reynolds Tobacco Company:

Jones, Day, Reavis & Pogue
Metropolitan Square
1450 G Street, N.W.
Washington, D.C. 20005-2088

BY: Peter J. Biersteker

ALSO PRESENT:

William J. Thompson, Research Analyst
Shook, Hardy & Bacon

Robert D. Scott, Attorney
Whyte Hirschboek Dudek

* * * *

I N D E X

EXAMINATION BY:	PAGE
MR. SILFEN	6

* * * * *

EXHIBIT INDEX

NUMBER		MARKED
2401	Expert Report of Drs. Zeger, Wyant and Miller (No Bates Numbers)	19
2402	Attachment B, Major Smoking Attributable Diseases (No Bates Numbers)	19
2403	ICD-9 Description (No Bates Numbers)	227

THE VIDEOGRAPHER: Good morning, we're on 08:38:06
the video record. Today's date is September 10, 08:38:08
1997. The time is now 8:38 a.m. 08:38:12
My name is Dave Jenkins, the video 08:38:14
technician, associated with Ray J. Lerschen's & 08:38:16
Associates. Today's witness is Scott Zeger. May we 08:38:20
have the introduction of counsel followed in by the 08:38:22
swearing of the witness. 08:38:24

MR. SILFEN: I'm Tom Silfen, counsel for 08:38:26
Philip Morris from Arnold & Porter. 08:38:30

MR. BIERSTEKER: Peter Biersteker from 08:38:34
Jones Day, counsel for R.J. Reynolds Tobacco 08:38:36
Company. 08:38:36

MS. STEURY: Ellen Steury from Arnold & 08:38:38
Porter, counsel for Philip Morris. 08:38:40

MR. SCHWARTZBAUER: Bob Schwartzbauer from 08:38:44
Dorsey & Whitney in Minneapolis, counsel for Philip 08:38:48
Morris Inc. 08:38:48

MR. HAMLIN: Thomas Hamlin, Robins, 08:38:50
Kaplan, Miller & Ciresi, counsel for Plaintiffs 08:38:52
State of Minnesota and Blue Cross/Blue Shield of 08:38:54
Minnesota. 08:39:08

SCOTT L. ZEGER, Ph.D.
the Witness in the above-entitled
matter after having been previously
duly sworn testifies and says as follows:

EXAMINATION

BY MR. SILFEN:

Q. Good morning, Dr. Zeger. I am Tom Silfen, and I am 08:39:14
counsel for Philip Morris, and this is the 08:39:16
deposition of Dr. Scott Zeger in the Minnesota AG 08:39:22
case, at least that's what I call it. 08:39:24

Now, Dr. Zeger, I'm going to start by 08:39:28
marking as an exhibit to the deposition the report 08:39:34
that I believe you authored, along with Dr. Wyant 08:39:40
and Dr. Miller. 08:39:40

We have a very exotic numbering system 08:39:46
here, so I'm not going to try to guess what the 08:39:48
number is. Does anyone know? 08:39:50

Why don't we proceed without it. I'm 08:40:04
going to mark, and when the time comes, as the first 08:40:08
exhibit this report which I will hand to the witness 08:40:10
and to Mr. Hamlin. 08:40:16

Now, what I've given you is the narrative 08:40:18
portion of your report and the footnotes. The whole 08:40:22
report and attachments have been designated and 08:40:26

identified in an earlier deposition. 08:40:28

I didn't want to pile it up here, but if 08:40:30
you want attachments at any time, they are here and 08:40:32
we probably will talk about them later. Okay? 08:40:38

It's better to answer audibly, even when 08:40:42
you would normally just shake your head. 08:40:44

A. Okay. 08:40:44

Q. Is this, at least on its face, your report? 08:40:54

A. Based upon this cursory examination, it appears to 08:40:58
be my report. 08:40:58

Q. If at any time it appears that it's something other 08:41:00
than your complete, narrative statement, then you 08:41:04
will let us know. 08:41:06

Now, we've previously had time to talk to 08:41:12
Dr. Wyant and Dr. Leonard Miller. 08:41:18
Have you had a chance to read their 08:41:20
depositions? 08:41:22
A. Yes. 08:41:22
Q. Okay. Have you read any other depositions taken in 08:41:30
these litigations in other cases? 08:41:32
A. Could you repeat that? I'm sorry, in other cases? 08:41:40
Q. Yes. 08:41:40
A. No. 08:41:42
Q. Have you read any depositions in this case other 08:41:44
than Doctors Wyant and Leonard Miller? 08:41:48

A. No. 08:41:48
Q. As you know then from reading the depositions, 08:41:54
Doctors Miller and Wyant indicated that there was 08:42:00
some division of responsibility for the report, 08:42:04
although they said it was a group project. 08:42:06
What did you consider to be your areas of 08:42:08
principal responsibility? 08:42:12
A. This was a collaborative effort. 08:42:14
Q. They said so and I'm not doubting that. 08:42:16
A. I was involved in all aspects. I didn't do any of 08:42:24
the computing calculations, myself, but was involved 08:42:28
in planning and designing the core model and in 08:42:34
reviewing and finalizing the refined model, and in 08:42:44
design and review of the work in the nursing home 08:42:48
model. 08:42:50
Q. Okay. And when you say refined model, you're 08:42:52

including what I think of as the diminished health 08:42:56
status model, I take it?

A. Yes. 08:42:58

Q. Maybe we should get our terms straight since I think 08:43:04
of the core model in the same terms, I believe, that 08:43:08
you do. 08:43:10

If I say refined disease model, will we 08:43:12
then be talking about the same thing, which is the 08:43:16
refined model with respect to lung cancer/COPD and 08:43:20

CHD/stroke? 08:43:22

A. Okay. 08:43:22

Q. And then I'll say the diminished health status model 08:43:26
when I mean the diminished health status model. 08:43:30

A. Okay. 08:43:30

Q. And I will say the nursing home model when I mean 08:43:32
the nursing home model. 08:43:34

A. Okay. 08:43:34

Q. Good. Do I take it then that I can generally expect 08:43:52
that you will be able to answer questions regarding 08:43:56
all aspects of the report? 08:43:58

A. I'll do my best. 08:44:04

Q. I understand. 08:44:04

A. I would say that as I did not actually sit at the 08:44:10
computer and do calculations either in the core 08:44:12
model or in the refined model in its aggregate, 08:44:18
there may be questions about the very specific 08:44:20
details which I would have to refer you elsewhere 08:44:28

http://legacy.library.ucsf.edu/tid/iqq07a00/pdf

on, but I'll do my best to answer all the 08:44:28
questions. 08:44:28

Q. I am probably going to go through the details as I 08:44:32
did with Dr. Miller. And if there are places 08:44:36
where -- obviously, if there are places where 08:44:40
Dr. Miller knows, just say so, and we will shortcut 08:44:44
the discussion. 08:44:44

With that in mind, did you encounter any 08:44:50
places where statements by Dr. Miller were not 08:44:58
completely correct to your view or did you approve 08:45:06
of everything he said? 08:45:06

A. Said where? 08:45:10

Q. In his deposition. 08:45:12

A. I read his deposition through rather quickly. It 08:45:20
was a lot of reading. And I didn't read it to sort 08:45:22
of think through every answer and decide whether I 08:45:24
agreed or disagreed. So, you know, I can't 08:45:28
really -- if you wanted to ask me a specific 08:45:32
example, I'd be happy to -- 08:45:32

Q. That's fair. 08:45:34

A. -- think it through. 08:45:34

Q. I was really wondering if you came armed with 08:45:36
anything in particular that -- 08:45:38

A. No. 08:45:38

Q. The same with Dr. Wyant, is there anything in 08:45:42
particular that he said that you know right now that 08:45:44
you have a different view on? 08:45:46

A. No. 08:45:46

Q. Okay. I also read Dr. Miller last night and had the 08:45:52
same experience. It's not easy reading. I say that 08:45:58
having asked a lot of the questions and I still 08:46:00
found it hard going. 08:46:00

I think that Dr. Miller said something to 08:46:10
the effect that you were principally involved in 08:46:14
defining the product. I wrote that down, those were 08:46:18
his words. 08:46:20
What's your reaction to that? Were you 08:46:24
principally involved in defining the product? 08:46:26

A. I'm not sure I know what that -- 08:46:28

Q. I'm not sure I know, either. 08:46:30

A. I would just go back to what I had said earlier as 08:46:34
to my role. I think that my previous answer 08:46:36
describes accurately what my role was. I was part 08:46:40
of a collaborative team that designed and 08:46:42
implemented these analyses. 08:46:44

Q. Okay. I think Dr. Miller also said that you were 08:46:48
principally responsible for organizing it and 08:46:52
essentially developing a way to communicate it 08:46:56
easily. 08:46:56
I took that to mean that you had had a 08:47:00
significant role in shaping the presentation. By 08:47:06
that I mean report; is that correct? 08:47:08

A. I had a significant role as did I think Doctors 08:47:14
Miller and Wyant. 08:47:20

Q. Did you draft it? 08:47:20

A. I drafted sections of it. 08:47:22

Q. Which sections did you draft? 08:47:24

A. Let's see, I think I drafted the section on -- I'm not sure if I drafted -- I can't remember if I drafted the text as it appears now. 08:47:30 08:47:34 08:47:36

I certainly sketched large sections of the core model, the ideas and the idea of a series of probabilities applied to the total, fractions applied to the total expenditures. 08:47:40 08:47:46 08:47:50 08:47:54

I think I drafted the section on goodness of fit. And there may be some other sections which don't come back to me immediately. 08:47:58 08:48:02 08:48:04

But I would say that each of us wrote part of it and certainly each of us carefully read and revised and worked together to get a final document. 08:48:08 08:48:12 08:48:14 08:48:14

Q. All right. Do you know Vincent Miller? 08:48:48

A. No. 08:48:48

Q. Do you know who he is? 08:48:50

A. I saw reference to Vincent Miller in one of the depositions. I don't know who he is other than the mention of the name. 08:48:58 08:49:00 08:49:02

Q. How about Jeffery Harris? 08:49:02

A. I know of an economist by the name of Jeffery Harris, who I believe is at MIT. I don't know if that's the Jeffery Harris to whom you're referring. 08:49:06 08:49:12 08:49:16

(Mr. Scott entered the deposition room.)

MR. HAMLIN: Do you know Mr. Scott? 08:49:20

MR. SILFEN: I don't.

MR. HAMLIN: Basic introductions here. 08:49:22
This is Bob Scott, he's one of the counsel for the 08:49:24
State of Wisconsin. He's also been admitted pro hac 08:49:28
vice here. This is Tom Silfen, Peter Biersteker -- 08:49:34

MS. STEURY: Ellen Steury. 08:49:34

MR. HAMLIN: I'm sorry, Ellen.

MR. SCOTT: I didn't mean to be a force of 08:49:38
test. 08:49:38

MR. THOMPSON: Tom Thompson. 08:49:40

MR. SCHWARTZBAUER: Bob Schwartzbauer. 08:49:42

MR. HAMLIN: This is Professor Scott 08:49:44
Zeger. 08:49:44

MR. SCOTT: Thank you. Sorry for the 08:49:48
interruption. 08:49:48

MR. SILFEN: Why don't we talk off the 08:49:52
record and tell me why your northern neighbor is 08:49:54
visiting. 08:49:54

(Discussion off the written record.)

BY MR. SILFEN:

Q. I take it you're not familiar with work that 08:50:14
Dr. Harris has done in these cases? 08:50:18

A. That's correct, I'm not familiar with his work. 08:50:20

Q. Are you aware that in this case the defense has 08:50:28

http://legacy.library.ucsf.edu/tid/iqq07a00/pdf

filed expert reports that relate to your report? 08:50:34
A. I am aware that expert reports have been filed. 08:50:38
Q. Have you read those reports? 08:50:42
A. I have skimmed through some of them. 08:50:44
Q. Now, is it your understanding that your group is in 08:50:56
some respects reworking its report in reaction to 08:51:04
some of the comments in the defense reports? 08:51:06
A. We're doing some additional calculations, most of 08:51:18
which represents a completion of work which was 08:51:22
undertaken and most of which is reported here in 08:51:26
this document. 08:51:30
I believe there may be one aspect of the 08:51:32
calculations that are being done partly at I think 08:51:40
my suggestion and the suggestion was also put 08:51:44
forward, I think, in one of those reports. 08:51:46
Although, I don't remember which one. 08:51:48
Q. Did you have a meeting with Doctors Wyant and Miller 08:51:56
to discuss what additional work you would do or not 08:52:00
do in reaction to those reports? 08:52:00
A. I recall a meeting where we discussed some aspects 08:52:10
of -- some comments made in those reports. I would 08:52:14
say the majority of the work that we were attempting 08:52:16
to complete now is independent of those comments and 08:52:22
is just trying to finish what was part of our 08:52:24
original plan. 08:52:28

Q. Well, if you read, and I know that you did read 08:52:30
Dr. Miller, you know that he, in fact, mentioned 08:52:32

that some work was being done in response to those 08:52:34
reports? 08:52:34

A. Yes, I can think of -- I mean, I don't remember his 08:52:38
specific comment, but I won't argue that he did say 08:52:42
that. I don't recall if he said that or not. 08:52:44

Q. Okay. When you had this meeting or conversation in 08:52:48
which the reports were discussed, did you and 08:52:56
Dr. Miller and Dr. Wyant prepare for the meeting by 08:53:04
reading the reports? 08:53:06

A. Actually, I did not read the reports in preparation 08:53:08
for that meeting. But there was some discussion of 08:53:12
the reports there. I had, prior to that meeting, 08:53:14
skimmed very quickly, ten minutes, for a stack of 08:53:18
documents that big. 08:53:22

But there was some discussion, and at that 08:53:24
report -- at that meeting there was some, you know, 08:53:26
a paragraph read here or there. 08:53:30

Q. Have you done more than skim them now or are you 08:53:34
still in the skim state? 08:53:34

A. That's where I am. 08:53:36

Q. Well, you're in the same state as Doctors Wyant and 08:53:42
Miller, you're all in the skim state. 08:53:44

Let's take a look at your report. 08:53:56

MR. SILFEN: Do we have an exhibit 08:54:00
number?

MR. SCHWARTZBAUER: No, we can't find the
keeper of the numbers.

BY MR. SILFEN:

Q. Well, we will proceed without a number. I will 08:54:06
refer to your report as your report and we will be 08:54:10
communicating. 08:54:10

A. So noted. 08:54:12

Q. On page 1, the first sentence, do you see that? 08:54:36

A. Yes. 08:54:40

Q. The first sentence states, I'll quote it, "We were 08:54:48
retained by the State of Minnesota (the state)," in 08:54:52
parens, "and Blue Cross Blue Shield of Minnesota, 08:54:56
(Blue Cross)," in parens, "to determine the amount 08:55:00
of money they expended in 1978-1996 to purchase 08:55:04
smoking attributable health care services." That's 08:55:08
the end of the quote. 08:55:08

I will from time to time quote material 08:55:16
that we both have in front of us and both can read. 08:55:20
That is being done strictly for the record so that 08:55:22
someone reading the record will know what we're 08:55:24
talking about. 08:55:24

A. I understand. 08:55:24

Q. It is a bother, but it's kind of the procedure. 08:55:30

Did you write that sentence? 08:55:34

A. I don't recall. 08:55:36

Q. Does this sentence define the task that you 08:55:42
understood you were undertaking? 08:55:46

A. Yes. 08:55:48

Q. And how was that task defined? How did you decide 08:55:56
this was the task? 08:55:58

A. Could you -- I don't understand the question. 08:56:00
Q. Well, I assume that when you set out on an 08:56:06
investigation, you first defined the task; isn't 08:56:10
that correct? 08:56:10
A. Yes, it is as stated here. 08:56:12
Q. Well, and how was it decided that that was the 08:56:16
task? 08:56:16
A. Well, this was what was asked of us by counsel. 08:56:22
Q. In other words, in these very words it was described 08:56:30
to you that this was the task? 08:56:32
A. No, these are our words, but we were asked to look 08:56:38
at the expenditures for this period of time that the 08:56:42
state or Blue Cross had made that were for health 08:56:46
care services attributable to smoking. 08:56:50
And we had a specific period of time and 08:56:52
that was the task we undertook. 08:56:56
Q. Do you recall any discussions about the definition 08:57:06
of the task and whether there were different tasks 08:57:14

that might have been or should have been 08:57:16
undertaken? 08:57:16
A. No. I was not party to a debate about or discussion 08:57:24
about whether this was the right task. This was the 08:57:28
task that we had been asked to undertake, so that's 08:57:30
what we focused on doing. 08:57:32
Q. Well, did there ever come a time when you questioned 08:57:48
whether another task or another definition of a task 08:57:52
would be appropriate? 08:57:54

A. No. 08:57:58

Q. I'm looking at the top of the second page, and in 08:58:50
the first sentence there's a reference to the major 08:58:56
smoking attributable diseases that are used in this 08:59:00
report. 08:59:00

A. Sir, did you say the second sentence? 08:59:02

Q. The first sentence. 08:59:04

A. Yes. 08:59:04

Q. A reference to the major smoking attributable 08:59:06
diseases that are used in this report. I am told 08:59:18
that the exhibit you have in front of you is number 08:59:22
2401. 08:59:26

A. I feel much better. 08:59:30

Q. I am greatly relieved, myself. 08:59:32

Now, I take it that from the other 08:59:44
depositions that the list of smoking attributable 08:59:50

diseases was given to you by Dr. Samet; is that 09:00:00
true?

A. Could I have that list? 09:00:02

Q. Sure, absolutely. Why don't we mark them both. 09:00:30
This is Attachment B. 09:00:40

MR. SILFEN: I'll ask the reporter to mark 09:00:40
the report as 2401 and Attachment B as 2402. 09:00:44

(Defendants' Exhibits 2401 and 2402 marked for 09:01:24
identification by the reporter.)

BY MR. SILFEN:

Q. Okay. I take it now we have marked the narrative 09:01:30
report as Exhibit 2401 and Exhibit B as Exhibit 09:01:36

2402. 09:01:40
My question referred to Exhibit B, which 09:01:42
the narrative describes as a definition of major 09:01:48
smoking attributable diseases; is that correct? 09:01:50

A. Attachment B, yes, is the list of major smoking 09:01:54
attributable diseases. 09:01:56

Q. And I understood from Dr. Miller and Dr. Wyant that 09:02:00
that list was given to you by Dr. Samet; is that 09:02:06
true? 09:02:06

A. That's my understanding, yes. 09:02:08

Q. So you had no part in deciding what the major 09:02:10
smoking attributable diseases were? 09:02:12

A. That's correct. 09:02:12

Q. The next sentence on page 2 states, and I'll quote, 09:02:26
"The state and Blue Cross paid for health care 09:02:30
services for more than 90,000 persons suffering from 09:02:34
these diseases during 1978 to 1996." 09:02:36
Do you see that? 09:02:38

A. Yes. 09:02:40

Q. Is it your understanding, as Dr. Miller told us, 09:02:46
that those are actually the 90,000 persons whose 09:02:58
experience is reflected in these reports? 09:03:02

A. I'm sorry, I don't understand the question. 09:03:06

Q. Well, these are the real people whose medical 09:03:10
experience is the basis for these reports, correct? 09:03:14

A. These are some of the people. 09:03:18

Q. Who are the other people? 09:03:22

A. Well, this report reflects expenditures for anybody 09:03:28
who received health care services in the period of 09:03:32
time from Medicaid, GAMC, or Blue Cross/Blue Shield, 09:03:40
of which these 90,000 are a part. 09:03:42

Q. Who are the other people that are not covered by the 09:03:48
state and Blue Cross? 09:03:48

A. Well, there might be people who are not suffering 09:03:52
from the specific diseases listed here that are also 09:03:56
a subject in this report. 09:03:58

Q. I see. 09:03:58

A. These are 90,000 -- well -- 09:04:02

Q. I see. So these would be the people in this report 09:04:08
suffering from the major smoking attributable 09:04:12
diseases? 09:04:12

A. That's my understanding. 09:04:14

Q. Okay. That's fair. That's good. And, in fact, had 09:04:26
I looked at the next sentence, that's the import of 09:04:30
the next sentence, I take it, in addition? 09:04:32

A. Correct. 09:04:32

Q. Okay. The middle paragraph on page 2 introduces a 09:04:42
chart indicating percentages of Minnesotans who had 09:04:50
a history of smoking. 09:04:52
Do you see that? 09:04:52

A. Yes. 09:04:52

Q. And what is the significance of that chart to this 09:04:56
report? 09:04:56

A. Could you -- I don't quite understand the question. 09:05:06

Q. You put it in the chart, you put it in the report, 09:05:10

and I want to know why it's here. What's its 09:05:12
importance? 09:05:12

A. This Table 1 lists the percentages of Minnesotans 09:05:18
age 19 and over who during the period '84 to '94 09:05:22
smoked or had a history of smoking. 09:05:26

And it shows that for persons covered by 09:05:32
their employer or union, about 25 percent are 09:05:36
current smokers and more than 50 percent are current 09:05:38

or former smokers, and that the rate is somewhat 09:05:42
higher for persons covered by Medicaid. 09:05:44

The significance is that a large fraction 09:05:46
of citizens in Minnesota or Minnesotans are either 09:05:52
current or former smokers. 09:05:54

Q. And in -- this is what we call in this report 09:05:56
sometimes ever-smokers, right? 09:05:58

A. When you say this is, which -- 09:06:02

Q. Current or former together are ever-smokers, 09:06:04
correct? 09:06:04

A. Yes, that's my understanding. 09:06:06

Q. And throughout this period, then, what this chart 09:06:10
tells us is that more than -- something more than 50 09:06:14
percent of Minnesotans age 19 and over were 09:06:18
ever-smokers? 09:06:20

A. Yes, that's correct. 09:06:22

Q. Let's take a look at paragraph 4 at the top of page 09:07:28
4. The first sentence says that, "The state and 09:07:36
Blue Cross paid substantially more nursing home 09:07:40

residence fees for smokers than for never-smokers." 09:07:44
Do you see that? 09:07:44

A. I do see it, yes. 09:07:46

Q. And Dr. Miller has already told us that the word 09:07:50
"Blue Cross" should not have been included, so 09:07:52
we'll just assume that that's out. Okay? 09:07:56

A. Okay. 09:07:56

Q. Do you agree with that? 09:07:58

A. That's my understanding. 09:08:00

Q. Okay. Is it your understanding also that that first 09:08:08
sentence relates to the finding of your nursing home 09:08:12
model? 09:08:14

A. Yes, that's my understanding. 09:08:16

Q. Is it your understanding, also, that that sentence 09:08:28
relates to the same period that we've talked about 09:08:34
in your opening sentence, 1978 to 1996? 09:08:38

A. Yes, that's my understanding. 09:08:44

Q. And so do I take it you, Dr. Zeger, are telling us 09:08:52
in this sentence that from 1978 to 1996 the state 09:08:56
paid more nursing residence fees for smokers than 09:09:04
for nonsmokers? Is that what you're telling me? 09:09:12

A. I'm sorry, could you repeat the question once 09:09:14
again? 09:09:16

Q. Are you telling us in this sentence, Dr. Zeger, that 09:09:22
during the period 1978 to 1996, that the state paid 09:09:36
more nursing home residence fees for smokers than 09:09:36
for never-smokers? 09:09:36

A. This sentence is intended to summarize the results 09:09:44

of the nursing home model. 09:09:46

Q. That's what you already told me. 09:09:50

A. The model was estimated from NHANES data for a 09:10:00

period 1982 to 1992, I believe, an 11-year period. 09:10:04
Various probabilities necessary to make a 09:10:08
calculation analogous to what we do for nonnursing 09:10:12
home expenditures were calculated from NHANES and 09:10:16
then applied to state expenditures for this period 09:10:20
1978 to 1996. 09:10:24

And based upon the calculations, the 09:10:28
specific calculations that were made, it was 09:10:32
determined that a certain fraction of the 09:10:36
expenditures actually made were attributable to 09:10:38
smoking. 09:10:42

I believe that's what this sentence is 09:10:44
intending to summarize. And in order to answer your 09:10:48
question specifically, I think it would be necessary 09:10:50
to, you know, look at the details there. 09:10:52

Q. I'm sure you will not be surprised to find that I do 09:10:56
not understand your answer. 09:10:58

Here's my question again: Is it your 09:11:00
testimony that during the period 1978 to 1996 that 09:11:08
the state paid substantially more nursing home 09:11:10
residence fees for smokers than for never-smokers, 09:11:16
is that your testimony? 09:11:16

A. That's what the sentence says. 09:11:18

Q. I want to know if it's also your testimony. Is that 09:11:20

what you're telling us here today? 09:11:20

A. I don't have any basis now to discount what this sentence says. 09:11:26 09:11:26

Q. That's kind of an unusual response. This is your report. 09:11:30 09:11:30

Is it your testimony that during the period 1978 to 1996, the state paid substantially more nursing home residence fees for smokers than for never-smokers? 09:11:34 09:11:40 09:11:44 09:11:44

A. I'm hesitating because this does not include a qualification as to the specific dates. 09:12:04 09:12:06

Q. You made the qualification that there was some extrapolation; subject to that qualification. 09:12:10 09:12:12

A. Yes. 09:12:12

Q. The next sentence states, I quote, "Smokers entering nursing homes during this period were far more likely than never-smokers to be suffering from lung cancer and chronic obstructive pulmonary disease." 09:12:24 09:12:26 09:12:30 09:12:34

Do you see that? 09:12:34

A. Yes. 09:12:36

Q. Are you familiar with the computation that supports that statement? 09:12:42 09:12:42

A. I'm generally familiar with it. 09:12:44

Q. What do you mean by generally familiar? 09:12:48

A. I had a discussion with Dr. Wyant and have briefly reviewed a table that he prepared from which this 09:12:56 09:13:00

sentence was drawn. 09:13:04

He was responsible for drafting this 09:13:06
sentence and preparing that table, but I have 09:13:12
briefly reviewed it. 09:13:12

Q. And does this also reflect your testimony, this 09:13:18
sentence? Do you believe this to be correct? 09:13:20

A. Yes. 09:13:20

Q. Did you review this table prior to the writing of 09:13:24
this report or did you review it after? 09:13:26

A. I think both. I think both times. I'm not -- I 09:13:30
don't recall specifically. 09:13:32

Q. Paragraph C-6 on page 4 you report the percentage of 09:14:08
smoking attributable expenditures for Blue Cross, 09:14:10
for GAMC and for Medicaid, do you see that? 09:14:14

A. Yes. 09:14:18

Q. And there is a difference between the percentages, 09:14:26
14.9 percent for GAMC, 10.7 percent for Blue Cross, 09:14:32
and 4.9 percent for Medicaid. 09:14:34

Do you see that? 09:14:34

A. Yes. 09:14:36

Q. Have you given any consideration to why the 09:14:40
attributable percentages should be so different? 09:14:48
For instance, why the GAMC percentage should be 09:14:54
three times the Medicaid percent? 09:14:56

A. I've given no specific consideration to that. 09:15:00

Q. What would be the explanation for such a result? 09:15:04

A. I can only speculate here. 09:15:08

Q. I understand. I'm just wondering what you would attribute it to, whether it's a concern? 09:15:12 09:15:14

A. It's not a concern. 09:15:16

Q. If it's not a concern, then you must have an explanation. What is the explanation? In other words, you're not startled and horrified? 09:15:18 09:15:24 09:15:26

A. Correct. 09:15:26

Q. So what is the explanation? 09:15:28

A. Well, these are different populations of people, they're different ages, they have different conditions, they're different in many ways. 09:15:32 09:15:34 09:15:40

And the whole idea of our approach to modeling was to allow the smoking attributable fractions to depend on things like what disease you have and how old you are and so forth. 09:15:42 09:15:46 09:15:50 09:15:54

And I'm assuming that these populations are different in those respects. 09:15:56 09:15:56

Q. So you're saying that it's -- the critical factor here is the different populations so that, for instance, you are not surprised to see a different smoking attributable fraction for a Medicaid population? 09:16:00 09:16:02 09:16:06 09:16:10 09:16:10

A. These are smoking attributable expenditures. 09:16:14

Q. Well, that's a fraction. Those aren't dollar numbers, those are fractions? 09:16:18 09:16:20

A. Yes, but it's the fraction of expenditures. 09:16:24

Q. All right, you're not surprised to see a 09:16:26
different --
A. That's correct, I'm not surprised. 09:16:26
Q. And that's because the Medicaid population is 09:16:28
different from the GAMC population? 09:16:36
A. I said it may be. You asked me to speculate; I 09:16:38
did. I gave you my best speculation. 09:16:38
Q. It was good speculation. 09:16:44
A. Thank you. 09:16:44
Q. On page 5, paragraph 9, there is discussion of the 09:17:06
confidence intervals. Is your work on confidence 09:17:16
intervals now complete? 09:17:18
A. I don't know. 09:17:18
Q. You're not doing it? 09:17:20
A. That's correct. 09:17:22
Q. Dr. Leonard Miller is doing it? 09:17:26
A. No. 09:17:26
Q. Dr. Wyant is doing it? 09:17:30
A. Yes. 09:17:30
Q. I was afraid you were going to say no, and then I 09:17:32
was going to not know who to turn to. 09:17:36
Now, I may have this wrong, but I believe 09:17:52

that Dr. Miller said that he had been doing the work 09:17:56
on confidence intervals, is that not right? 09:18:00
A. I don't know what he said. I don't remember. 09:18:04
Q. Well, but who historically has been doing the work 09:18:06
on confidence intervals? 09:18:08

A. Well, the three of us met and discussed the approach 09:18:14
to calculating confidence levels and agreed upon an 09:18:20
approach. And Dr. Wyant and Dr. Miller have been 09:18:26
doing the computations all along. 09:18:28

I believe Dr. Wyant is doing the 09:18:30
computations of confidence intervals for the core 09:18:34
model. And I think Dr. Wyant and Dr. Miller working 09:18:40
together are doing the computations for the refined 09:18:42
model. 09:18:44

But I think it would be fair to say that 09:18:46
Dr. Wyant is sort of trying to get this thing 09:18:48
finished up, leading the effort. 09:18:50

We all, of course, participated in the 09:18:52
design of the approach to calculating confidence 09:18:54
levels. 09:18:56

Q. Is your report significant? 09:18:56

A. I don't know what you mean. 09:19:00

Q. Well, is it -- does it have a confidence interval 09:19:04
that passes through zero or 1 or whatever -- 09:19:12

A. The results are not completed yet. 09:19:12

Q. So you do not know at this day whether all or parts 09:19:16
of your report pass a test of significance? 09:19:18

A. I don't know whether they do; I don't know whether 09:19:22
they don't. I don't have any information about 09:19:24
that, so I can't answer that question. 09:19:24

Q. I understand. Well, suppose a part didn't pass a 09:19:36
significance test. You're doing a jackknife, is 09:19:38
that correct, or that's what I heard originally? 09:19:40

A. We're using the statistical method of jackknifing, 09:19:44
yes, to make confidence testing. 09:19:46

Q. I won't dare ask you what that is, but let me leap 09:19:50
to the bottom line. 09:19:50
Suppose that the core model doesn't pass 09:19:52
the test, that the confidence interval includes one, 09:20:00
if that's the proper terminology. Would that be the 09:20:02
proper terminology? What would flunking be? 09:20:06
Describe to me flunking. 09:20:10

A. We're not doing any test, so there's no passing or 09:20:14
failing. 09:20:14

Q. You're not doing a test? 09:20:16

A. That's correct. 09:20:16

Q. So we won't come out with something comparable to a 09:20:22
passes significance at a 5 percent level? 09:20:24

A. We're calculating a confidence interval, which is to 09:20:30
say that we will report an estimate of the smoking 09:20:34

attributable expenditures for the period 1978 to 09:20:38
1996. 09:20:38
And then we will give an interval of 09:20:40
values that reflect the uncertainty in our overall 09:20:44
estimate. 09:20:44

Q. But you will nowhere produce what I think of as a 09:20:50
standard measure of significance? Won't that be 09:20:56
something that falls out of that project? 09:21:00

A. It is possible in calculating a confidence interval 09:21:04
to perform a statistical test by asking whether a 09:21:10

particular value falls within or outside of that 09:21:12
confidence interval. 09:21:16
Q. Yes. 09:21:16
A. But our plan right now is to calculate that 09:21:20
confidence interval. And I don't have anything else 09:21:24
to, you know, to say about whether any particular 09:21:26
test will be done or not. 09:21:28
Q. Well, was it ever your plan to do tests of 09:21:30
significance? 09:21:30
A. Tests of significance have been used in some of the 09:21:34
work that's reported here, but it's not been my 09:21:38
intention to test any particular hypothesis. 09:21:44
Q. So you are not going to report the results of those 09:21:48
tests of significance? 09:21:50
MR. HAMLIN: Objection; asked and 09:21:52

answered. 09:21:54
MR. SILFEN: I don't think so, Tom. I 09:21:56
mean, if you're making a work product objection -- 09:21:56
MR. HAMLIN: Asked and answered 09:21:58
objection. 09:21:58
MR. SILFEN: I didn't know that I had 09:22:00
the -- unless the answer is no, I perceived the 09:22:02
answer to be no. 09:22:04
BY MR. SILFEN:
Q. I perceive you to be saying you are not going to 09:22:06
report a test of significance; is that correct? 09:22:08
A. That's not what I said. 09:22:10
Q. Well, then what did you say? 09:22:12

A. What I said is that we are going to report a 09:22:14
confidence interval, which is an expression of the 09:22:16
uncertainty in the point estimate of the smoking 09:22:20
attributable expenditures. 09:22:22

And then I said that it is possible, 09:22:24
having received, having obtained a confidence 09:22:26
interval, for the reader to conduct their own test 09:22:28
or for the writer to have a test which says, "Does a 09:22:32
particular value fall within or outside that 09:22:36
interval?" 09:22:36

And that would be a valid test of 09:22:38
hypothesis. But what we plan to report -- and what 09:22:50

we plan to report is a confidence interval for the 09:22:50
smoking attributable expenditures. 09:22:52

Q. And why wouldn't you also take the next step and 09:22:54
report on the significance of the result? 09:22:56

A. The word "significance" is a vague word and it has 09:23:04
many different meanings to different people. You 09:23:08
haven't specified a particular hypothesis. 09:23:12

There are many hypotheses, I mean, that 09:23:14
one might test and report a significance for. So is 09:23:18
there a specific hypothesis that you're referring 09:23:24
to? 09:23:24

Q. In paragraph 11 at the bottom of page 5, you talk 09:24:22
about adjustments that are made for characteristics 09:24:28
of smokers. Do you see that? 09:24:30

A. Paragraph 11 on the bottom of page 5? 09:24:32

Q. Yes. 09:24:34

A. Yes. 09:24:34

Q. And you say at the end of that paragraph that it is 09:24:38
unlikely that any further adjustments would 09:24:44
materially change our smoking attributable 09:24:46
expenditure estimates? 09:24:48

A. Yes. 09:24:50

Q. Is that your view? 09:24:52

A. Yes. 09:24:52

Q. And what is the basis for that view? 09:24:54

A. The Surgeon General's Report in 1989 is one basis 09:25:00
for it. 09:25:00

Q. Explain to me why. 09:25:02

A. That report has a sentence very much to this effect 09:25:08
saying that even though there's been an effort to 09:25:12
look at other explanations for the smoking effect on 09:25:20
disease and smoking-caused diseases there has been 09:25:26
very little change in the estimated smoking effects 09:25:28
from those efforts to do additional control. 09:25:32

Q. Do you have any other support for that theory, for 09:25:36
this sentence? 09:25:38

A. Yes. In addition to the 1989 Surgeon General's 09:25:40
Report, there are a few other pieces of evidence 09:25:44
which we base this on. 09:25:48

One is the book by Will Manning in which 09:25:56
he did have available to him information on alcohol, 09:26:02
the data sets he used, and estimated his models, his 09:26:06
medical expenditures, controlling for all of the 09:26:12

variables he had, including alcohol and exercise, I 09:26:16
believe, and then not controlling for any of them. 09:26:18

And his result changed, I think, by 10 or 09:26:22
15 percent. And that's when he excluded all of the 09:26:28
variables, such as socioeconomic status and 09:26:30
everything, changed about 15 percent. So that's the 09:26:34
second piece of evidence. 09:26:36

Q. And is there a third? 09:26:38

A. Yes, there is a third. The similarity in our 09:26:42
results from the core model in which we didn't 09:26:46
control for factors like socioeconomic status and 09:26:50
education and the refined model where we did control 09:26:54
for those factors were quite similar to one 09:26:58
another. And that's a third piece of information. 09:27:02

Q. Do you have another basis or do we have them all 09:27:10
now? 09:27:10

A. Let's see. Yes, there is one more that comes to 09:27:14
mind. There's a paper published from the Framingham 09:27:20
group that I recall reading, let's see if I can 09:27:26
recollect, I think it's in the "American Journal of 09:27:30
Medicine" where they looked at the Medicare 09:27:32
expenditures for their Framingham population and 09:27:40
looked at a series of risk factors for 09:27:44
cardiovascular disease, of which smoking was a major 09:27:46
one. 09:27:48

There were three, I believe, smoking being 09:27:48
a major one. And estimated the health care, 09:27:54

estimated the smoking effect or these cardiovascular 09:27:56
risk effect on cardiovascular-related expenditures, 09:28:00
and then controlled, in addition to that, for 09:28:02
exercise and alcohol and other things and found none 09:28:06
of those variables to affect the result. 09:28:08

Q. Have you done this research, yourself, to establish 09:29:22
these four factors? 09:29:26

MR. HAMLIN: Objection to form. 09:29:30

THE WITNESS: I'm not sure I understand 09:29:32
the question. 09:29:32

BY MR. SILFEN:

Q. You gave me four reasons for the view expressed in 09:29:40
paragraph 11. 09:29:40

And what I'm asking you: Is this the 09:29:42
result of your research, or is this an answer to 09:29:52
this question that has been supplied to you by other 09:29:54
members of your group? 09:29:58

A. No, it was my answer to your question. Some of that 09:30:04
work which I've just described was done by me in 09:30:06
collaboration with Doctors Wyant and Miller, that 09:30:12
would be the calculation of the core model and the 09:30:14
refined model. 09:30:16

The other three citations were citations 09:30:22
to previous research, not to my research. 09:30:26

Q. I understand that. But that's previous research 09:30:28
that you're familiar with personally and you have 09:30:32
read? 09:30:32

A. Yes. 09:30:32

Q. And you've read the 1989 Surgeon General's Report or 09:30:36
the pertinent parts? 09:30:38

A. Yes. 09:30:38

Q. And the statement you're referring to is from 09:30:42
chapter 3, is it not? 09:30:42

A. I don't recall the chapter. 09:30:44

Q. Did the work in chapter 3 control for anything at 09:30:48
all? 09:30:48

A. I don't recall the chapter, so I don't know what's 09:30:52
in chapter 3. 09:30:54

Q. Did the work in the Surgeon General's Report on the 09:30:56
risks of smoking control for anything at all, 09:30:58
anything? 09:30:58

A. I don't recall the specifics reported there. The 09:31:04
sentence I cited was a statement about the effort to 09:31:12
control. As everything in the Surgeon General's 09:31:16
Report, it was a reference to work in the 09:31:16
literature, not to their own work. 09:31:18

Q. Can you recall what work was referenced in that 09:31:20
sentence? 09:31:20

A. I do not. 09:31:22

Q. Did you ever look? Did you ever try and find the 09:31:26
work that's referenced in that sentence? 09:31:28

A. All I recall is the sentence at this point. 09:31:30

Q. So the answer is you don't recall ever looking to 09:31:32
see if there was anything actually referenced? 09:31:34

A. From that particular sentence? 09:31:36

Q. Yes. 09:31:36

A. No. 09:31:38

Q. Had you read Dr. Manning's work apart from your -- 09:31:56
this task here in Minnesota, or did you read it in 09:32:00
connection with this Minnesota work? 09:32:02

A. I read it in connection with this work. 09:32:06

Q. So you're familiar with Manning's methodology? 09:32:10

A. It's been a while, but I'm somewhat familiar. 09:32:14

Q. Did he define the task the same way that you have? 09:32:24

A. What task are you referring to? 09:32:28

Q. The task that you have defined in the first sentence 09:32:34
of your report. 09:32:36

A. I don't think his task was to establish the smoking 09:32:42
attributable expenditures for the State of 09:32:46
Minnesota, no. 09:32:46

Q. Well, forget for a moment the State of Minnesota. 09:32:54
Was his task to establish the smoking attributable 09:32:56
expenditures for the population generally? 09:33:00

A. I believe the -- no, I would say no. 09:33:08

Q. What was his task? 09:33:10

A. I believe his main goal was to look at questions of 09:33:12
excise taxes and the relationship of taxes to 09:33:20
internal and external costs, largely external costs, 09:33:26
of smoking. 09:33:28

Q. I'm not sure that's a distinction. I mean, you're 09:33:32

right, he was trying to see if smoking pays its way, 09:33:36
but in order to do that, of course, he had to figure 09:33:38
out what smoking costs, right? 09:33:40

A. He looked at a -- yes. 09:33:50

Q. And the fact that he was looking at external costs 09:33:54
doesn't distinguish your work here because you're 09:33:56
also looking at external cost, right? 09:33:58

A. I don't recall his specific definition of external 09:34:02
and internal, so I'm not sure. 09:34:02

Q. Yes, but the fact is you're looking at external 09:34:06
costs, too? 09:34:06

A. If you define "external costs" for me, I could 09:34:08
answer that question. 09:34:10

Q. You don't know what external costs are? You used 09:34:14
the term first. 09:34:14

A. I was trying to refer to the word "external costs" 09:34:20
as described by Manning, and I don't recall his 09:34:24
specific definition at this time. 09:34:26

Q. Well, he surely included costs that were borne by an 09:34:36
insured group and not by the individual who was 09:34:42
sick, let's say? That's clearly part of his 09:34:44
external costs, isn't it? 09:34:46

A. The costs borne by an external group were certainly 09:34:50
part of his external costs? 09:34:52

Q. Yes. 09:34:54

A. To the best of my recollection, yes. 09:34:54

Q. That's what you've been measuring here, right? 09:34:58

A. We tried to estimate the fraction of the dollars 09:35:00
expended by the State of Minnesota that's 09:35:04
attributable to smoking. 09:35:04

Q. Tell me what's the difference in his task and yours, 09:35:08
as he defined it? 09:35:10

A. I don't remember his specific task as he defined 09:35:12
it. I've given you a sense of involving something 09:35:20
about excise taxes. I don't remember the specifics 09:35:22
as he's formulated it. 09:35:24

Q. Okay. On page 6 in paragraphs 14 and 15 you say 09:35:52
that your analysis goes beyond methodological and 09:35:58
data limitations of previous analyses. Do you see 09:36:02
that? 09:36:02

A. Let's see, 14 is our analysis goes beyond major 09:36:06
methodological limitations, and 15 beyond the major 09:36:12
data limitations, yes. 09:36:14

Q. What particular prior analyses did you have in 09:36:18
mind? 09:36:20

A. One prior analysis is the model used and reported in 09:36:32
the MMWR that was done by, I believe, Bartlett and 09:36:42
Miller and others. And I believe that's one of the 09:36:46
prior works that's being described here. 09:36:50

Q. Any other? 09:36:54

A. I think there's probably also reference to a 09:36:58
previous work done for the Centers for Disease 09:37:04
Control, also which I think it goes under the name 09:37:08
SAMMEC. 09:37:08

Q. Why is your work superior to SAMMEC? 09:37:22

A. To SAMMEC, sir? 09:37:26

Q. Yes. 09:37:28

A. This is to the best of my understanding. SAMMEC was 09:37:38
developed using indicators of resource utilization, 09:37:46
such as hospital visits and the like, as opposed to 09:37:50
specific expenditure data. 09:37:54
I think it was based upon the I think the 09:38:00
1989 health interview survey. Whereas we were able 09:38:06
to actually use real expenditures for actual 09:38:08
diseases. That's to the best of my recollection. 09:38:14

Q. SAMMEC, does SAMMEC control for anything other 09:38:26
than age and gender? 09:38:28

A. I don't recall what else they control for. 09:38:30

Q. Are you familiar with SAMMEC because of your work 09:38:36
for Minnesota or because it's just something you're 09:38:40
familiar with in the field? 09:38:40

A. It's through my work on this project. 09:38:52

Q. You had, however, done some smoking-related studies 09:38:52
in the past, hadn't you? 09:38:52

A. One. 09:38:54

Q. What was that one? 09:38:54

A. I actually was involved in a study looking at the 09:39:00
price elasticity of tobacco use using state data. 09:39:12
It was actually a project done by student 09:39:14
-- the masters of public health students at Johns 09:39:16
Hopkins. And I was, again, involved in helping in 09:39:20

this study. 09:39:22
Q. That's the only tobacco-related work you had done? 09:39:32
A. I can't recall any other specific examples. 09:39:34
Q. I thought I saw some others, but we'll take a look 09:39:38
later. 09:39:38
MR. HAMLIN: Tom, when you get to a point, 09:39:46
I'd like a break.
MR. SILFEN: This is fine, we can take a 09:39:50
break now. 09:39:50
THE VIDEOGRAPHER: We're temporarily going 09:39:52
off the video record. The time is now 9:39 a.m. 09:39:56
(Off the record.) 09:39:58
THE VIDEOGRAPHER: We're back on the video 09:46:04
record. The time is now 9:46 a.m. 09:46:06
BY MR. SILFEN:
Q. All right, continuing the deposition. You mentioned 09:46:12
the model that was published in the MMWR, correct, 09:46:20
the Miller/Bartlett model? 09:46:26
A. Yes. 09:46:26

Q. Are you familiar with that model? 09:46:28
A. Not very familiar with it. 09:46:28
Q. When did you start, as best you recall, working on 09:46:50
this Minnesota Project? 09:46:52
A. October -- I think end of October 1996. 09:46:56
Q. When you arrived, did you then have input in the 09:47:10
shaping of the project? I'm thinking particularly 09:47:14
of your statement earlier this morning that you had 09:47:20
a major role in deciding on the core approach, the 09:47:24

three reductions? 09:47:26
A. Did I have -- I'm sorry, repeat the question. 09:47:30
Q. I guess what I'm saying is almost tautological. You 09:47:36
told us that you had a role in deciding on the three 09:47:38
reductions methodology, correct? 09:47:40
A. Yes, I had a role in formulating the core model, 09:47:46
which involves three reductions. 09:47:48
Q. But the three reductions are used in, as you related 09:47:54
in this report, in all your models; isn't that 09:47:56
true? 09:47:56
A. Yes. 09:47:58
Q. So this core structure of three reductions flows 09:48:02
through the whole report, does it not? 09:48:04
A. Correct. 09:48:04
Q. Now, I take it, then, that that three reductions 09:48:10
structure was not in place when you arrived in 09:48:14

October of 1996; is that true? 09:48:16
A. No. 09:48:18
Q. It was in place? 09:48:20
A. Aspects of it were in place. 09:48:22
Q. Well, what was it that you contributed then? 09:48:30
A. I contributed a number of things, including 09:48:34
proposing that we do a core simplified analysis, 09:48:42
making the three reductions explicit. 09:48:48
Those reductions are inherent in the 09:48:50
refined model, although, not made explicit in the 09:48:58
way they are during the core model, at least they 09:49:04

were inherent in the calculations that were being 09:49:06
done when I arrived, you know, the work on the model 09:49:10
when I arrived. 09:49:10
But one of my contributions was to help 09:49:12
make explicit their use. 09:49:14
Q. And what were your other contributions? 09:49:20
A. You know, we have worked as a team since that date. 09:49:28
All decisions made about all models are really based 09:49:32
upon conversations that we had with each other. 09:49:34
Q. When you arrived, was the model disease specific, as 09:49:46
it is today? 09:49:46
A. The refined model? 09:49:48
Q. Any part of the work that was ongoing. 09:49:50
A. The work on the refined model, I believe, was 09:49:54

disease specific by November when I arrived. It's 09:50:00
my understanding that Dr. Samet played a large role, 09:50:06
and he had arrived, I don't know how much earlier, 09:50:08
but somewhat earlier than I had. 09:50:10
Q. I see. Dr. Samet is also at Johns Hopkins, 09:50:16
correct? 09:50:18
A. Yes. 09:50:18
Q. Have you worked with him previously? 09:50:22
A. Yes. 09:50:24
Q. On what projects? 09:50:26
A. We collaborate on a project looking at the mortality 09:50:34
effects of particulate air pollution. 09:50:38
Q. I take it you mean this is an ongoing work that you 09:50:42
do with him? 09:50:42

A. That's correct. 09:50:42
Q. Is that the only project you have shared with him? 09:50:46
A. To the best of my recollection, yes. 09:50:50
Q. Who is that work being done for? 09:50:52
A. The work is currently funded by the Health Effects 09:50:56
Institute in Boston. 09:51:00
Q. Okay. All right. Let's take a look at the three 09:51:06
reductions, and I am going to look at page 7, and I 09:51:10
am going to also look at the same time at pages 9 09:51:16
and 10 where you have an example. 09:51:20
Are you familiar with that example? 09:51:22

A. Yes. 09:51:22
Q. Did you construct it? 09:51:24
A. With Doctors Wyant and Miller. 09:51:32
Q. I have this picture of the three of you kind of 09:51:34
huddled around the table? 09:51:36
A. We're very close friends. 09:51:38
Q. Good to hear that. All right. 09:51:42
The first reduction, "How many smokers?" 09:51:54
I take it this is take into account the simple fact 09:51:58
that some of the disease occurs among nonsmokers, 09:52:06
and the disease incurred by nonsmokers is not part 09:52:10
of the smoking attributable cost? 09:52:12
A. That's correct. 09:52:14
Q. Okay. And so in the simple example that you have on 09:52:30
page 9, there are 160 total lung cancers, 20 of them 09:52:30
are among never-smokers so that you, in the first 09:52:36

step, would reduce the attributable dollars by the 09:52:44
20, or a proportion constructed of the 20? 09:52:50

A. In the first step we would reduce the expenditures 09:52:58
by the fraction .875, which would be 140 divided by 09:53:02
160, leaving out the 20 nonsmokers. 09:53:04

Q. Thank you. That's a better way to say it. 09:53:06
Now, the second reduction is titled, "How 09:53:22
much extra disease?" Do you see that on page 7? 09:53:26

A. Yes. 09:53:26

Q. What do you mean by how much extra disease? 09:53:30

A. Well, what that refers to is that even in a 09:53:42
population of smokers, even in a population -- even 09:53:52
with people who have lung cancer among smokers, that 09:53:54
we would expect some of them to have had lung cancer 09:54:00
-- well, we would expect a certain baseline rate of 09:54:04
lung cancer among an otherwise similar population of 09:54:08
nonsmokers. 09:54:08
So we don't want to attribute all of the 09:54:12
lung cancers among the smokers to their smoking, as 09:54:16
we recognize that a certain fraction occur even 09:54:20
among similar nonsmokers. 09:54:22
So this second reduction is attempting to 09:54:24
set aside those. 09:54:24

Q. And that baseline is the amount that you assume the 09:54:30
smoking population would have occurred even had they 09:54:38
never smoked, as you put it here; is that correct? 09:54:40

A. I would word it slightly differently. I would say 09:54:44
it's the increased rate of lung cancer in the 09:54:50

smoking population over what would occur in an 09:54:54
otherwise similar nonsmoking population. 09:54:58

Some people like to talk about this 09:55:02
hypothetical group not smoking. It's simpler for me 09:55:08
to think about an otherwise similar nonsmoking 09:55:12
populations. 09:55:12

Q. What did you mean when you said "baseline"? 09:55:16

A. You'd have to -- 09:55:20

Q. You used the word "baseline." 09:55:22

A. I don't remember how I used it. 09:55:24

Q. What do you mean -- you said a baseline rate, what 09:55:28
do you mean by that? 09:55:28

A. What I mean is the rate that would occur in an 09:55:30
otherwise similar group of nonsmokers. 09:55:32

Q. Are you familiar with the term expected in 09:55:34
epidemiology, the expected rate, and the observed 09:55:40
rate, do you know what those are? 09:55:40

A. Not as you're using them, no. 09:55:44

Q. Well, would the baseline rate be the rate that you 09:55:48
would expect in a population of smokers if they 09:55:52
weren't smokers? 09:55:54

A. I'm sorry, could you repeat that? 09:55:58

Q. Would the baseline rate be the rate of lung cancer 09:56:02
that you would expect in a population of smokers if 09:56:08
they weren't smokers? 09:56:10

A. That's not the way I use the word baseline, no. 09:56:12

Q. So you -- do you have a different view? I mean, 09:56:16

http://legacy.library.ucsf.edu/tid/iqq07a00/pdf

read a sentence from page 7 here, it's the second 09:56:20
full sentence, "We calculate how often the smokers 09:56:22
being treated got the disease above and beyond the 09:56:26
rate at which they would have gotten it anyway, had 09:56:30

they never smoked." 09:56:30
A. I was clarifying the way I would put that idea. You 09:56:40
know, it's a semantic, it's somewhat semantic the 09:56:44
difference. I'm not going to -- well, I would tend 09:56:52
to write it in the way that I stated it, comparing 09:56:56
otherwise similar populations. 09:56:58
This is another way some would write it. 09:57:00
And obviously the person who wrote it here wrote it 09:57:02
that way. I don't remember who wrote that 09:57:04
sentence. 09:57:06
Q. Well, but I understood that you were principally 09:57:08
involved with the core model and in particular with 09:57:12
these reduction steps; isn't that true? 09:57:14
A. Yes. 09:57:14
Q. So I take it you closely reviewed this 09:57:16
presentation? 09:57:18
A. Yes. 09:57:20
Q. In fact, Dr. Miller said you were principally 09:57:24
involved with the communication in the presentation 09:57:26
of just this, these steps; isn't that correct? 09:57:30
A. I don't know what he said. I don't recall what he 09:57:32
said. 09:57:32
Q. Look at the final sentence of the paragraph, "We 09:57:36
then reduce the total expenditures a second time to 09:57:38

remove the expenditures for diseases that would have 09:57:42

occurred anyway." Do you see that? 09:57:44
A. Yes, I do. 09:57:46
Q. Do you agree with that? 09:57:46
A. Yes. 09:57:46
Q. You said a moment ago that the distinction you were 09:58:12
drawing was semantic, correct? 09:58:16
A. I don't recall exactly what I said, but something to 09:58:20
that effect, yes. 09:58:22
Q. Is the distinction you're drawing anything other 09:58:24
than semantic, or is it just a semantic difference? 09:58:28
A. It can be more than semantic in some applications, 09:58:40
but I would have to look at a specific question. 09:58:44
Q. Well, what did you have in mind? 09:58:46
A. I don't know what you mean by the question. 09:58:48
Q. You said it could make a difference in some 09:58:54
applications, and I said what do you have in mind? 09:58:56
A. Well, what I had in mind was that if you were to ask 09:59:00
me a specific -- give me a specific example of where 09:59:04
this, you know, and ask me whether this was more 09:59:06
than a semantic difference in that example, I would 09:59:10
be happy to try to answer. 09:59:14
Q. Well, I assumed that you meant that you could think 09:59:14
of circumstances in which it would be more than 09:59:16
semantic. That's my question, Dr. Zeger. 09:59:20
A. Yes, I could think of circumstances. 09:59:22

Q. And what are they? 09:59:22
A. There may well be circumstances where one might 09:59:28
imagine the world as it was and how it might have 09:59:32
been where it's so difficult to formulate how it 09:59:38
might have been that it becomes very important to 09:59:42
make the distinction between describing the 09:59:46
difference between how the world was and how it was 09:59:48
in a similar group of people who didn't smoke versus 09:59:52
how it might have in some hypothetical situation. 09:59:58
Q. You better explain that further. I'm not sure what 10:00:00
you're talking about. 10:00:02
You're saying there are some circumstances 10:00:04
where it might be very difficult. What 10:00:04
circumstance? What do you have in mind? You 10:00:08
obviously have something in mind here, tell us. 10:00:10
A. I don't have anything else in mind, no. I was 10:00:12
trying to be responsive to your question is there 10:00:14
ever more than a semantic difference, and I said 10:00:16
yes. 10:00:18
Q. You said something about a -- well, let's read back 10:00:20
the answer, the witness's answer. 10:00:22
(The requested portion read back.) 10:01:08
BY MR. SILFEN:
Q. So you said there may well be circumstances where 10:01:10
one might imagine the world as it was and how it 10:01:12

might have been where it's so difficult to formulate 10:01:14
how it might have been, it becomes very important to 10:01:18
make the distinction? 10:01:18

A. To make the semantic -- I said that the distinction 10:01:22
would be more than semantic in that case is what I 10:01:24
was trying to say. 10:01:26

Q. What case is that? You said one might imagine, what 10:01:28
do you imagine? 10:01:28

A. I could imagine a situation where you talked about 10:01:34
the difference between intelligence math ability for 10:01:40
boys and girls, and you could imagine comparing 10:01:46
otherwise like boys and girls, or you can imagine 10:01:48
having those boys go back and become girls, and then 10:01:52
think of the world that way. 10:01:54

But that seems more than a semantic 10:01:56
difference because I don't know how to make boys 10:01:58
into girls. 10:01:58

Q. Do you know how to make smokers into nonsmokers? 10:02:02

A. I'm not really the best person to talk about that. 10:02:06
I would say Dr. Samet might be better than me. 10:02:10

Q. I guess I'm not -- I'm missing something here. On 10:02:14
page 7, the construct that you've created, the 10:02:20
second reduction says, "How much extra disease?" 10:02:24

A. Right. 10:02:24

Q. And it must be -- if it's extra disease, it must be 10:02:32

excess disease, right, excess disease? 10:02:36

A. Okay. 10:02:36

Q. And that means that it is disease that would not 10:02:40
have occurred in some other circumstance, doesn't 10:02:44
it? It's got to be excess over something. 10:02:48

A. What I've tried to say in this line of questioning, 10:02:54
my answers to your questions, is that what we have 10:02:58
done is to calculate the rate of cancer among 10:03:04
smokers as compared to the rate excess over the rate 10:03:08
of cancer among otherwise similar nonsmokers. 10:03:14
That's what I mean by excess. 10:03:16

Q. And so what you have done is you have defined the 10:03:20
excess as the difference between the smoker's 10:03:24
experience and the experience of a nonsmoker, 10:03:32
holding some other factors equal, correct? 10:03:34

A. Not of a nonsmoker, no, of a population of 10:03:38
nonsmokers. 10:03:40

Q. Right, you've done that on a population basis? 10:03:42

A. Yes, thank you. 10:03:42

Q. So you know how to compare smokers and nonsmokers 10:03:48
and figure the difference, don't you? That's what 10:03:50
you've done? 10:03:50

A. That's what we've done, yes. 10:03:52

Q. And for purposes of that calculation, you assumed, 10:04:30
did you not, that had the smokers not smoked, they 10:04:34

would have had the disease incidence rate of the 10:04:40
similar nonsmokers in the population? 10:04:44

A. What calculation are you referring to? 10:04:50

Q. This calculation, the excess. 10:04:52

A. Table 2? 10:04:52

Q. Reduction 2. 10:04:52
A. No, that assumption is not necessary to this 10:04:56
calculation, no. 10:04:56
Q. It isn't? 10:05:00
A. No. 10:05:00
Q. Explain to me why not. 10:05:02
A. Well, what we've done in Table 2 is to compare the 10:05:08
rate -- 10:05:08
Q. I'm talking about how much extra disease 10:05:12
calculation, the way you've just described it. 10:05:14
For purposes of that reduction step, isn't 10:05:18
it true that you have assumed that had smokers, had 10:05:24
these people who smoked not smoked, they would have 10:05:28
had the disease incidence experience of people who 10:05:32
never smoked? 10:05:34
MR. HAMLIN: Objection; asked and 10:05:34
answered. 10:05:36
THE WITNESS: What I'm telling you is the 10:05:42
calculation that we did is not based upon that 10:05:44
assumption. The calculation is based upon -- well, 10:05:48

that's what I said before. 10:05:50
BY MR. SILFEN:
Q. Well, let's look at the calculation there. I take 10:06:18
it that what you do in step 2 is you see that the 10:06:24
smokers experienced 140 lung cancers, correct? 10:06:32
A. In Table 2, the smokers had 140 lung cancers, that's 10:06:36
correct. 10:06:36

Q. And how much lung cancers were extra? 10:06:40

A. Well, what we do -- you're talking about in 10:06:42
reduction 2? 10:06:44

Q. Yes. 10:06:44

A. What we do is we look at the rate of lung cancer 10:06:48
among an otherwise similar group of nonsmokers. 10:06:54
Here 20 out of 5,000. And we take as extra the 10:07:00
difference between the rate for the smokers relative 10:07:02
to the rate for the never-smokers. 10:07:04

Q. And so the smokers experienced 140, the 10:07:10
never-smokers experienced 20, and how many are 10:07:14
excess or extra? 10:07:16

A. Well, let's see, it says there are 120 attributable 10:07:26
lung cancers out of 140. And the 120 is arrived at 10:07:30
by noticing that there are 20 out of 5,000 in the 10:07:36
never-smoker group. 10:07:38

And if you apply that, if you say how much 10:07:40
higher is the rate for the smokers than for the 10:07:44

never-smokers, you would get a reduction that gets 10:07:46
you down from 140 to 120. 10:07:48

Q. I understand. But isn't what you're saying that 20 10:07:54
of the cases among the smokers would have been 10:08:00
expected to occur even if these people had not 10:08:04
smoked? 10:08:04

A. Say it one more time. I'm sorry, I don't mean to 10:08:14
split hairs with you. 10:08:14

MR. SILFEN: Why don't you read it back. 10:08:16

(The requested portion read back.) 10:08:26

THE WITNESS: So long as you use the word 10:08:30
"expectation" in a precise statistical sense, I 10:08:34
would agree with that. 10:08:36

BY MR. SILFEN:

Q. And that's fine, but what do you mean by in a 10:08:44
precise statistical sense? 10:08:44

A. Well, the word "expectation" means on average over, 10:08:48
you know, population of possible events and 10:08:52
basically would get me back to a model that compares 10:08:54
the smokers to the never-smokers. 10:08:56

Q. Very good. 10:08:56

A. Population of smokers to population of 10:09:00
never-smokers. 10:09:00

Q. Very good. I'm tempted to point out to you that at 10:09:16
the beginning of the discussion I asked if we could 10:09:18

do this in terms of expected and observed and you 10:09:22
seemed to not want to do it that way. We muddled 10:09:26
around for a bit. 10:09:26

A. I apologize for muddling around. And you said 10:09:30
expected and observed. In the epidemiological 10:09:34
sense, epidemiologists use those words very loosely 10:09:38
sometimes, and I wasn't willing to do so. 10:09:38

Q. But now we've used them to your satisfaction? 10:09:42

A. Yes. 10:09:42

Q. Now, on page 10 you say at the top, second sentence 10:10:10
of the first full paragraph, that the two reductions 10:10:16
that we've just discussed is an example of the 10:10:24

calculation of attributable risk, a standard 10:10:26
approach in epidemiology. Do you see that? 10:10:30
A. Yes, I do. 10:10:30
Q. In fact, aren't these two steps simply what's 10:10:38
commonly known as the attributable risk formula? 10:10:42
A. Yes. 10:10:42
Q. Okay. And that goes way back in time to an article 10:10:48
by Levin, or are you familiar with that? 10:10:52
A. I'm familiar with the reference to Levin. I've not 10:11:00
read his article recently or ever. 10:11:02
Q. I first saw it in a book by the Lillienfelds. Have 10:11:04
I got the name right? 10:11:06
A. Lillienfeld, yes. 10:11:08

Q. When I was trying to understand this stuff. And the 10:11:14
Lillienfelds refers to the attributable risk formula 10:11:20
as the maximum percentage that could be 10:11:24
attributable. 10:11:26
Do you know why they referred to it that 10:11:26
way? 10:11:28
A. I don't know specifically why they referred to it 10:11:32
that way, no. 10:11:32
Q. And I think, just to make sure we're talking about 10:11:44
the same thing, I think of the attributable risk 10:11:46
formula as being the prevalence of the exposure 10:11:52
times the quantity relative risk minus one over the 10:11:56
prevalence of the exposure times quantity R minus 10:12:00
one plus one. Is that familiar to you? 10:12:02
A. It's familiar to me, yes. 10:12:06

Q. And is that what you meant, as well, when you said 10:12:14
this was the standard attributable risk formula? 10:12:18

A. I didn't say it was the standard attributable risk 10:12:22
formula. I said -- the application of these 10:12:26
reductions is an example of the calculation of 10:12:28
attributable risk. 10:12:30

Q. I intended to ask a more precise question. In fact, 10:12:32
these two steps taken together are the attributable 10:12:34
risk formula, which I just related, they are 10:12:38
mathematically the same thing; isn't that correct? 10:12:40

A. That's correct. 10:12:40

Q. And isn't it also correct that that's SAMMEC? Isn't 10:12:54
that just SAMMEC? 10:12:54

A. SAMMEC uses attributable risk. 10:13:00

Q. That same formula? 10:13:02

A. SAMMEC uses the attributable risk in its method, 10:13:06
yes. 10:13:06

Q. And when you say attributable risk, that is the same 10:13:10
formula we just discussed, right? 10:13:12

A. SAMMEC uses that formula for the attributable risk 10:13:14
in its calculations. 10:13:16

Q. So so far have we done anything other than what 10:13:26
SAMMEC does? 10:13:26

A. When you say so far, you're referring to? 10:13:32

Q. These first two steps, you haven't controlled for 10:13:36
anything, you've just applied the attributable risk 10:13:40
formula. 10:13:40

A. So far we've not talked about any application of 10:13:42
these ideas. So far we have a hypothetical example 10:13:44
and we're trying to illustrate the two reductions 10:13:48
which you and I have talked about which comprise 10:13:58
attributable risk. 10:13:58

Q. Right. In the hypothetical example we've controlled 10:13:58
for no other factors and we've just applied the 10:14:00
attributable risk formula. That's all we've done, 10:14:00

right?

A. Correct. 10:14:00

Q. That is also the way the core model works, isn't it, 10:14:04
for the first two reduction steps? You control for 10:14:06
nothing but age and gender and you apply the 10:14:10
attributable risk formula? 10:14:12

A. Age and gender and insurance status. 10:14:16

Q. Okay. 10:14:18

A. And disease, disease group. 10:14:30

Q. As an outcome, disease is your -- 10:14:32

A. Well, we do the calculation separately for lung 10:14:36
cancer and COPD, and we do it again separately for 10:14:40
the other major smoking attributable diseases. 10:14:42

Q. And in these first two steps, are you modeling 10:14:46
expenditure? 10:14:48

A. No. 10:14:48

Q. You're modeling incidence, correct? 10:14:54

A. No. 10:14:54

Q. What would you call it? 10:14:56

A. What would I call what? 10:15:00

Q. What is it you're modeling if it's not expenditure 10:15:04
and it's not disease incidence, what is it? 10:15:06

A. Well, in these first two steps we're estimating the 10:15:08
probability of smoking, given you have a particular 10:15:12
disease, as well as the probability of having that 10:15:18

disease, having that disease given you're a smoker 10:15:22
and given you're a nonsmoker. Those are the things 10:15:24
we're calculating in these two steps.

Q. The two steps taken together gets you an 10:15:28
attributable fraction, correct? 10:15:30

A. Gets you an attributable fraction, that's correct. 10:15:34

Q. And that attributable fraction is a fraction based 10:15:40
on comparative expenditure between smokers and 10:15:44
nonsmoker? 10:15:44

A. No. 10:15:44

Q. It is based on comparative what? 10:15:48

A. Rates of disease. 10:15:50

Q. And is that different to your view than incidence of 10:15:54
disease? 10:15:54

A. Yes. 10:15:54

Q. What is the difference between the term rates of 10:15:58
disease and incidence of disease? 10:16:00

A. The word "incidence" has a -- has units of inverse 10:16:06
time. It's the rate occurrence of new disease per 10:16:12
unit time, per unit time, whereas we're not making 10:16:18
that calculation here. Calculations we're using are 10:16:24
more commonly called prevalence, not incidence. 10:16:26

Q. I see. Actually, I -- you are correct, and I was 10:16:30
just thinking of incidence and the common sense of 10:16:34
an event. 10:16:36

But what you're saying is that the term 10:16:38
incidence to you is a time based calculation which 10:16:44
is opposed to prevalence; that's just what you 10:16:50
said? 10:16:50

A. Yes. 10:16:52

Q. And what you're doing here is a prevalence 10:16:56
computation? 10:16:56

A. Correct. 10:16:58

Q. Is that true of all aspects of the model, of the 10:17:02
report, is it all prevalence? 10:17:04

A. Yes. To the best of my recollection at the moment 10:17:14
of all the different things we do. 10:17:16

Q. I think that's -- well, I won't answer the 10:17:26
question. I'll stick with your answer. 10:17:28

Maybe you should define some terms for me 10:17:34
so that we're talking the same language. 10:17:42
Cross-sectional/longitudinal, do they have different 10:17:52
meanings than prevalence/incidence? 10:17:54

A. Yes. 10:17:56

Q. Okay. Can you describe for me the difference 10:18:02
between the two pairs of terms? 10:18:04

A. I could give you a brief definition of each of the 10:18:12
terms, if that would start us down the road. 10:18:14

Q. You could also do a long lecture, I'm sure, about 10:18:16
that, but do the brief description. 10:18:18

A. A cross-sectional study is one where you look at the 10:18:22
occurrence of disease or study of disease. You look 10:18:26
at the occurrence of disease at a fixed point in 10:18:30
time. In the longitudinal study, one follows people 10:18:34
over time and observes the occurrence of disease 10:18:38
over time. 10:18:42
Cross-sectional studies are typically used 10:18:44
to estimate the prevalence of disease. Longitudinal 10:18:50
studies can be used to either estimate the 10:18:52
prevalence of disease or the incidence of disease, 10:18:54
and are often used to do both. 10:18:58

Q. What about a cohort study? Where would that fit 10:19:06
into all those terms? 10:19:08

A. A cohort study is an example of a longitudinal study 10:19:12
where a population of people is followed through 10:19:14
time. 10:19:14

Q. Are there particular problems for which one or the 10:19:22
other of those types of studies are appropriate? 10:19:26

A. As I said, it's possible to estimate prevalence from 10:19:38
a longitudinal study or from a cross-sectional 10:19:40
study. 10:19:42
So if you had a problem where you wanted 10:19:44
to estimate prevalence, you could use either sort of 10:19:46
study. 10:19:48
If you wanted to estimate incidence, you 10:19:50

wouldn't be able to estimate incidence without 10:19:54
auxiliary information from a cross-sectional study, 10:19:56
and so in that case a longitudinal study would be 10:20:00
necessary. 10:20:00

Q. And that would also be a cohort study? 10:20:02

A. You could do so from a cohort study, yes. 10:20:08

Q. And that was helpful, but what I was actually 10:20:12
looking for is: Are there types of problems for 10:20:16
which a longitudinal or incidence approach is more 10:20:22
appropriate and types of problems for which a 10:20:26
cross-sectional or prevalence approach is more 10:20:30
appropriate? 10:20:30

MR. HAMLIN: Objection to form. 10:20:34

THE WITNESS: Yeah, I've not made this 10:20:36
pairing of incidence and longitudinal and prevalence 10:20:38
and cross-sectional. 10:20:38

BY MR. SILFEN:

Q. I will unpair them if that will help. 10:20:42

A. Why don't you ask it again. 10:20:42

Q. Are there particular -- and if I choose the wrong 10:20:46
pair, just tell me. We started with incidence and 10:20:50
prevalence, so I'll go back to them. 10:20:52

Are there particular kinds of problems for 10:20:54
which an incidence study is more appropriate and 10:20:58
particular kinds of problems for which a prevalence 10:21:00

study is more appropriate? 10:21:02

MR. HAMLIN: Objection to form. 10:21:06
THE WITNESS: Well, incidence, the 10:21:08
incidence of disease is a quantity that you may be 10:21:10
interested in. If you're interested in incidence, 10:21:14
then you should do a study that enables you to 10:21:16
estimate incidence. 10:21:18
If you're interested in prevalence, you 10:21:20
should do a study that enables you to estimate 10:21:22
prevalence. 10:21:22
For example, you can do a cross-sectional 10:21:24
study or a longitudinal study. If you're interested 10:21:28
in incidence, you should do a study that enables you 10:21:30
to estimate that. So it depends what the scientific 10:21:32
question is. 10:21:32

BY MR. SILFEN:

Q. That's fair. But what I was getting at is whether 10:21:36
there are particular kinds of questions which are 10:21:40
ordinarily approached through incidence or more 10:21:44
readily approached through incidence and other kinds 10:21:48
of problems that are more readily approached through 10:21:50
prevalence? 10:21:52

MR. HAMLIN: Objection to form. 10:21:54

THE WITNESS: Yes. 10:21:58

BY MR. SILFEN:

Q. Could you tell me what they are? 10:22:00

A. Yes. If you want to know the incidence of disease, 10:22:06
then you should do a study that enables you to 10:22:10

estimate incidence. 10:22:10

I mean, if your question is what is the 10:22:12
incidence of lung cancer, then you should do an 10:22:16
incidence study. 10:22:18

Q. What is that? I think what you're saying is 10:22:20
probably more enlightening to you than it is to me. 10:22:22
To me it sounds like the same thing over. 10:22:24

I was actually looking for an example of a 10:22:26
type of problem to which you would kind of 10:22:30
instinctively say, well, I'll do an incidence 10:22:32
study. 10:22:34

But what do you mean when you say if 10:22:36
you're interested in the incidence of lung cancer, 10:22:38
what does that mean? If I'm interested in the 10:22:40
incidence of lung cancer, what is it I'm interested 10:22:42
in? Tell me. 10:22:42

MR. HAMLIN: Objection to form. 10:22:44

THE WITNESS: Well, you're interested in 10:22:48
the rate of occurrence of new cases of lung cancer 10:22:52
per unit time, per unit time per person. 10:22:56

BY MR. SILFEN:

Q. The rate of occurrence of new cases. And describe 10:23:08

the other side of the coin then. Prevalence I would 10:23:12
be interested in what? 10:23:12

A. In the fraction of persons that have lung cancer at 10:23:18
a given time. 10:23:18

Q. And that would refer to a past exposure? It's the 10:23:24
burden of past accrual of the disease at the present 10:23:28

point in time? 10:23:30
A. Well, it's the fraction of people at the present 10:23:34
point in time who have that disease. Different 10:23:38
diseases have different mechanisms. 10:23:38

But it would be, you know, the number of 10:23:42
people per population unit who have the disease at a 10:23:44
particular time. 10:23:46

Q. Okay. Now, back to my original question. Can you 10:23:50
give me examples of problems, for instance, problems 10:23:54
you've addressed, that called for an incidence 10:24:04
approach? 10:24:04

A. I'm trying to think of what I've done in my career. 10:24:12
It's sort of murky right now. 10:24:16

Q. I share with you. I'm feeling that my career is 10:24:20
also murky right now. We're in the same murky 10:24:24
boat. We can commiserate at lunch. 10:24:26

A. I can give an example where prevalence is the 10:24:34
focus. Would that be helpful? 10:24:38

Q. I'd really like an example of both. Let me -- I 10:24:44

take it this is an area of expertise for you; is it 10:24:46
not? 10:24:46

A. I wouldn't consider myself, you know, a person with 10:24:54
special expertise in this area. These are basic 10:24:56
terms used in epidemiology. I am a biostatistician 10:25:00
who is knowledgeable about epidemiology. 10:25:04

Q. I didn't -- you are the -- what is your title? 10:25:10
You're the head of the department of epidemiology at 10:25:12

Johns Hopkins? 10:25:12

A. No, I'm not. 10:25:14

Q. What are you? 10:25:14

A. I'm a professor and chair of the department of 10:25:16
biostatistics at Johns Hopkins University. 10:25:18

Q. Is there a separate department of epidemiology? 10:25:22

A. Yes. Dr. John Samet is the chairman of 10:25:26
epidemiology. 10:25:28

Q. I'm sorry. 10:25:28

A. No problem. 10:25:30

Q. Yet unless I was also mixing the C.V.s, I believe 10:25:42
you've written a book on longitudinal analysis, 10:25:46
haven't you? 10:25:46

A. I've coauthored a book entitled, "The Analysis of 10:25:50
Longitudinal Data," yes. 10:25:52

Q. So unless it's a very, very, very short book, you 10:25:56
must have some familiarity with circumstances in 10:25:58

which longitudinal data is analyzed and 10:26:02
appropriate? 10:26:02

A. I'm sorry, is analyzed appropriately you said? 10:26:10

Q. Yes. 10:26:10

A. Yes. 10:26:10

Q. Now, why don't you give us some examples of when 10:26:14
longitudinal data has been used and when it has been 10:26:16
used in an incidence study. 10:26:18

A. Well, let's see, I've been involved with a study 10:26:26
looking at children's morbidity, children in 10:26:34
Indonesia. And you might be interested in the 10:26:36

question, what is the incidence rate for respiratory 10:26:42
infection among those children? 10:26:44

And you might be interested in asking 10:26:44
whether vitamin A deficiency predisposes children to 10:26:50
have an increased incidence of respiratory 10:26:54
infection. 10:26:54

Or you might ask a different question of 10:26:58
those same data. You might ask the question, what 10:27:00
is the prevalence of respiratory infection? And 10:27:04
does vitamin A deficiency predispose children to 10:27:08
have a higher prevalence of respiratory infection? 10:27:12

Q. And, in fact, that's the study referenced in the 10:27:20
section you wrote on goodness of fit? 10:27:22

A. It's actually a different study, but a similar 10:27:24

problem, yes. 10:27:26

Q. So I may just end up repeating what you just said, 10:27:46
which means I will probably garble it, but I want to 10:27:50
get it straight. 10:27:52

You said that if you were interested in 10:27:56
the occurrence of respiratory disease among persons 10:28:02
with a vitamin A deficiency, that a prevalence 10:28:10
study, a snapshot at a point in time, would be 10:28:14
appropriate. Did I say that correct? 10:28:14

A. I didn't say that. 10:28:16

Q. I understand. I was trying to get -- am I correct? 10:28:20

A. I don't know how to answer the question. I mean, 10:28:26
are you -- prevalence study, what I said is a 10:28:28

prevalence study may be of interest. 10:28:32
You might want to know are vitamin A 10:28:34
deficient children more likely to have at any given 10:28:40
time or averaged over a period of time more likely 10:28:42
to have respiratory infection than children who are 10:28:46
not vitamin A deficient? 10:28:48
That's a different question. That's a 10:28:50
prevalence question. It can be studied in a 10:28:52
longitudinal analysis, as is illustrated in the book 10:28:56
you referred to. 10:28:58
Or you could ask a different question. 10:29:00
You could ask, is the occurrence of a new case of 10:29:02

respiratory infection, is that influenced by whether 10:29:06
or not the child is vitamin A deficient? And those 10:29:10
are different questions and refer to different 10:29:12
aspects of the biologic process by which a child has 10:29:16
respiratory infection. In order to have it, you can 10:29:20
get it but then you have to keep it. 10:29:22
So prevalence reflects both the incidence 10:29:26
of the event but also the duration of time for which 10:29:30
it occurs. 10:29:30
Q. Now, if you -- in the study you reference in your 10:29:36
article, the one that I'm familiar with, the 10:29:48
children or some group of them were receiving 10:29:52
vitamin A supplements; is that correct? 10:29:54
A. There is a paper -- yes, in this report we referred 10:30:02
to a paper by Somer and Zeger in which the example 10:30:08
was a supplementation trial, yes. 10:30:12

Q. And would that have involved longitudinal data? 10:30:18

A. The data that are reported there, to the best of my 10:30:28
recollection, were collected -- well, they were 10:30:30
definitely collected in the longitudinal study. But 10:30:34
the data that are reported there would not be an 10:30:36
example of what I would call longitudinal data. 10:30:40

Q. Well, I'm less interested in that. If I wanted to 10:30:42
know, if I administer vitamin A to a group of 10:30:50
children and compare them to a group that does not 10:31:00

get vitamin A supplement -- 10:31:00

A. Right. 10:31:02

Q. -- and I want to know the effect of that experiment, 10:31:10
what analysis would I use? 10:31:12

A. It would depend on what the question you're asking 10:31:18
is. What specific question do you want to ask of 10:31:22
those data? 10:31:22

Q. Well, I'm sorry if I'm at a loss. I would assume 10:31:28
that the question would be, does the administration 10:31:34
of vitamin A supplements change or affect the 10:31:42
occurrence of whatever condition you're studying? 10:31:46

A. Well, I'm asking because I need to know what 10:31:48
condition. 10:31:50

Q. Oh, what is the condition you were studying at the 10:31:52
time? Was it mortality? 10:31:52

A. Yes. 10:31:54

Q. Okay. Then let's say mortality. 10:31:56

A. Okay. Got you. So we're going to look at the 10:31:58

mortality and the vitamin A supplementation. 10:32:00

Q. Yes. 10:32:02

A. Then if you'll just repeat your question. 10:32:04

Q. Whether vitamin A supplements have a healthy or an unhealthy or no impact on mortality? 10:32:12 10:32:14

A. Okay. 10:32:18

Q. And how would you do that study? 10:32:20

A. Well, the way that study was done is children were followed forward for a fixed period of time. I believe it was a year. 10:32:26 10:32:28 10:32:30

At the beginning of the period they were randomized to receive either vitamin A or not. And the number of children that died in the interval -- and death is an incident event, not a -- well, it's prevalent in the sense that you're gone, but it's different from respiratory infection in that it only happens once. 10:32:32 10:32:36 10:32:40 10:32:44 10:32:52 10:32:52 10:32:54

So the number that died in the group that got vitamin A were compared to the number that died in the group that didn't. 10:32:56 10:32:56 10:33:00

Q. Am I to understand that is an incidence approach? I say that only because you used the word incidence just now. 10:33:04 10:33:06 10:33:08

A. The question of interest was about the rate of mortality, which is an incident event, yes. 10:33:12 10:33:14

Q. I was trying to translate it into the terms you used before. If I were interested in -- suppose that we were considering death from malnutrition. I'm 10:33:18 10:33:22 10:33:28

picking that out of the air. 10:33:30
We would be looking at the new cases of 10:33:34
malnutrition among the persons who got the 10:33:36
supplement and the persons who didn't get the 10:33:40

supplement, correct? 10:33:40

A. Death or malnutrition? 10:33:44

Q. Malnutrition. 10:33:46

A. So we're now switching away from death and looking 10:33:48
at -- 10:33:48

Q. Yes. 10:33:50

A. -- some measure of malnutrition. 10:33:50

Q. I was trying to hook up with your term "new cases" 10:33:54
which made sense to me. I understood when we were 10:33:56
describing incidence, I understood that you were 10:33:58
saying that there you're looking at the new cases, 10:34:06
and I was trying to find an example that illustrated 10:34:08
that. 10:34:08
And would that be an example if the 10:34:10
outcome we were interested in was malnutrition and 10:34:14
we gave one group a vitamin A supplement and another 10:34:16
group not and we compared the new cases of 10:34:20
malnutrition in the group that got the supplement 10:34:22
and the group that didn't? 10:34:24

A. That would be an example of an incidence study, yes, 10:34:26
a study whose purpose was to estimate the incidence, 10:34:30
the relative incidence rates for vitamin A deficient 10:34:34
versus vitamin A sufficient children. 10:34:36

Q. If the question were, as it was there, whether the vitamin A supplement reduces the occurrence of 10:34:50 10:34:56

malnutrition, would it have to be done by an incidence study? 10:35:02 10:35:04

A. No. 10:35:04

Q. How could you answer that question without an incidence study? 10:35:06 10:35:08

A. I could do a cross-sectional study. I could randomize children to receive vitamin A or not, and then six months, two months later come back and do a cross-sectional study and see what the rate, the prevalence was of malnutrition in the two groups. 10:35:12 10:35:16 10:35:20 10:35:24 10:35:28

Q. I see. And that -- you referred to death as an incidence event. Why do you call that an incidence event as opposed to malnutrition? 10:35:40 10:35:44 10:35:46

A. Well, death happens once per person, at least so far as I know. 10:35:52 10:35:52

Q. We may all be surprised. 10:35:56

A. Yes. Malnutrition can come and go, respiratory infection can come and go. 10:36:00 10:36:02

Q. That's very interesting, and thank you for your patience. 10:36:16 10:36:16

Now, where we were when we got on to incidence and prevalence was we were up to the third step or the third reduction step. And that is described on the top of page 8. 10:36:28 10:36:30 10:36:42 10:36:50

Here's my question: We've already 10:36:54

determined the standard smoking attributable 10:36:58
fraction, or as you put it on page 10 the 10:37:14
attributable risk. 10:37:20
What more is there to do? Why do we need 10:37:24
a third step? 10:37:24

A. Well, it's likely that persons with major smoking 10:37:36
attributable diseases will have larger expenditures 10:37:42
than persons without -- than otherwise similar 10:37:46
persons without those diseases. 10:37:48
But it is our expectation that there is 10:37:56
some baseline level of expenditure in that otherwise 10:38:00
similar population. And, therefore, that we 10:38:04
shouldn't attribute all of the expenditures for a 10:38:08
person, for a group of people who are smokers, to 10:38:12
their smoking since otherwise similar groups also 10:38:14
have some expenditures. 10:38:16
So we take only the difference, the amount 10:38:18
by which the expenditures are increased in the 10:38:22
smoking group, and create a fraction which we apply 10:38:26
as the third reduction. 10:38:26

Q. Well, I take it that the point is that even if the 10:38:42
people have not smoked and not gotten a 10:38:48
smoking-related disease, you would expect them to 10:38:50
have some baseline cost? 10:38:54

A. That's correct. 10:38:54

Q. Well, isn't what you're doing then trying to 10:39:32
identify the dollar expenditure that would have 10:39:40
existed had this person never smoked? Isn't this 10:39:56
just a part of the same analysis? 10:39:58

A. No. 10:39:58

Q. Well, I'm -- would you agree with me if I said 10:40:08
you're trying to identify the dollar expenditure 10:40:10
that would exist for a similar nonsmoking person? 10:40:18

A. No. 10:40:20

Q. Well, then I am confused. In this paragraph at the 10:40:28
top of page 8, the first sentence says, "The smokers 10:40:34
receiving medical services for diseases attributable 10:40:36
to their smoking would, on average, have incurred 10:40:38
some health care cost anyway, even if they had not 10:40:42
gotten a smoking attributable disease." 10:40:44

A. Right. 10:40:46

Q. So aren't you trying to calculate the costs that 10:40:58
would have been incurred by a person who is a 10:41:16
never-smoker without the disease? 10:41:18

A. No. 10:41:18

Q. Well, Dr. Zeger, take a look at your next sentence. 10:41:28

A. The one that says we calculate how much greater -- 10:41:32

Q. -- the average expenditures are for smokers with a 10:41:34
smoking attributable disease relative to average 10:41:38
expenditures of never-smokers without the disease. 10:41:40

A. That's what we do. 10:41:42

Q. And the extra part -- well, but aren't you 10:41:56

identifying a baseline dollar amount that would be 10:41:58
expended by a never-smoker without the disease? 10:42:02
A. No. 10:42:02
Q. Okay. Well, you mean you do not identify -- let's 10:42:08
take this part. You do identify the average 10:42:18
expenditures of never-smokers without the disease, 10:42:20
don't you? 10:42:20
A. We estimate the average expenditures for, I'm sorry, 10:42:24
never-smokers without the disease? Yes, we do do 10:42:28
that, for a population of such people, yes. 10:42:30
Q. Is your problem with my articulation that I say 10:42:34
person rather than group? 10:42:34
A. That's one problem I have, yes. I wouldn't know how 10:42:38
to speculate about a particular person's 10:42:40
expenditures. 10:42:42
Q. I didn't mean to -- well, then let me say it again. 10:42:50
On a population-wide average basis, aren't you 10:42:54
trying to determine here the expenditure that would 10:43:04
have occurred for a never-smoker without a major 10:43:12
smoking-related disease, so that that amount will 10:43:16
not be attributed to smoking? 10:43:18
A. I would answer no again. And just so it's 10:43:24
absolutely clear what we do do, we calculate how 10:43:28

much greater the average expenditures are for 10:43:30
smokers with a smoking attributable disease than for 10:43:34
the average expenditures for never-smokers without 10:43:38
the disease. 10:43:38

That's what we attempt to estimate in this 10:43:42
calculation. We take that in and put it into a 10:43:44
fraction and apply that fraction to the total 10:43:46
expenditures for the population. 10:43:48

Q. Yeah, but doesn't that mean that you are removing 10:43:50
from the smoking attributable amount the fraction 10:43:54
that is represented by the average expenditures of 10:43:58
never-smokers without the disease? 10:44:00

MR. HAMLIN: Objection to form. 10:44:04

THE WITNESS: It just went on too long, 10:44:04
I'm sorry, I lost track. 10:44:06

MR. SILFEN: We'll read it back. 10:44:30

(The requested portion read back.)

MR. HAMLIN: Same objection.

THE WITNESS: You'll have to read it 10:44:32
again, I'm sorry, I don't think it was a complete 10:44:34
statement of what we're doing. 10:44:34

(The requested portion read back.) 10:44:48

MR. HAMLIN: Same objection. 10:44:50

THE WITNESS: We're not removing from the 10:44:52
smoking attributable amount anything. We're trying 10:44:54

to calculate the smoking attributable expenditures. 10:44:56

BY MR. SILFEN:

Q. This is called a reduction step, Dr. Zeger. I mean, 10:45:00
are you telling me that your third reduction step 10:45:02
does not remove an amount? 10:45:04

A. Not from the smoking attributable expenditures, it 10:45:06
doesn't remove an amount, no. 10:45:08

Q. What does it remove it from? 10:45:08
A. It ends up getting us the smoking attributable 10:45:12
expenditures. 10:45:12
Q. So it isn't a reduction? 10:45:14
A. It's a reduction, yes, it is a reduction. 10:45:16
Q. So it's a reduction. And isn't it a reduction which 10:45:20
is measured by -- well, let's go back to the 10:45:30
beginning. 10:45:30
Let's say in the population if you're a 10:45:40
smoker and you get lung cancer, you spend $1,000 for 10:45:48
the treatment of that lung cancer. You with me? 10:45:52
A. Well, go on, yeah. 10:45:58
Q. And let's say we find -- let me suppose that a 10:46:08
nonsmoker, an average nonsmoker without lung cancer 10:46:12
or any other smoking-related disease, on the average 10:46:16
spends $200 on his medical conditions, correct? You 10:46:26
with me? 10:46:26
A. I hear you. 10:46:30

THE VIDEOGRAPHER: We've got to cut. 10:46:32
MR. SILFEN: Fine, you can think about 10:46:34
it. The rest of the question is: We don't want to 10:46:36
charge smoking with the rest of the $200, that is 10:46:40
the question, and that is what this is about. Okay, 10:46:42
break time. 10:46:44
(A break was taken.) 10:46:52
THE VIDEOGRAPHER: We're back on the video 10:54:18
record. This is the second tape in the videotape 10:54:22

testimony of Scott Zeger. The time is now 10:54 10:54:24
a.m. 10:54:26

BY MR. SILFEN:

Q. Dr. Zeger, we've been talking about the three 10:54:32
reduction steps of the core model, correct? 10:54:36

A. Correct. 10:54:36

Q. I want to talk for a moment about the dollar figure 10:54:42
that those reduction steps are applied to. 10:54:46

A. Okay. 10:54:48

Q. And I'm going to call it the pot. 10:54:50

A. Okay. 10:54:52

Q. Because that's what Peter and I have been saying for 10:54:58
months. That's the only way we know how to say it. 10:55:00
If you look at page 22 -- 10:55:02

A. Page 22? 10:55:06

Q. Yes. 10:55:08

A. Yes. 10:55:10

Q. You have in your four indented paragraphs, you have 10:55:12
the -- well, paragraphs 2, 3 and 4 are the three 10:55:18
reduction steps. But what comes before that is what 10:55:26
I'm calling the pot that the reduction steps are 10:55:28
applied to. 10:55:28
And that is, quote, "Total expenditures 10:55:30
for persons with smoking attributable disease"; is 10:55:34
that correct?

A. Correct. 10:55:36

Q. Now, when we say "total expenditures for persons 10:55:40
with smoking attributable disease," I take it we do 10:55:42

not mean only the expenditures for that smoking 10:55:48
attributable disease?

A. Correct. 10:55:52

Q. We mean any expenditures during the period under 10:55:56
study by a person with a major smoking attributable 10:56:04
disease?

A. It's the expenditures for that person with a major 10:56:10
smoking attributable disease in a one-year period. 10:56:12

Q. And in this case, we're using the term person 10:56:16
because we really are talking about individuals? 10:56:18

A. Well, we're talking about all of the individuals who 10:56:24
have a major smoking attributable disease, yes. 10:56:28

Q. And what that means, for instance, is that if, say, 10:56:38

you're a person with lung cancer or COPD, you with 10:56:42
me so far? 10:56:42

A. Yes. 10:56:42

Q. You are then going to be treated in your lung 10:56:48
cancer/COPD model, correct? In the core or in the 10:56:58
refined disease? 10:56:58

A. You will be included in the expenditures, the total 10:57:06
expenditures, for persons who have lung cancer or 10:57:08
COPD, correct. 10:57:10

Q. And, for instance, in the refined disease model, 10:57:16
when we do a calculation to figure the smoking 10:57:20
attributable risk, that is a calculation that is 10:57:22
specific to the lung cancer/COPD persons, correct? 10:57:26

A. Correct. 10:57:28

Q. Now, if I understood your prior answer, the dollars 10:57:32
that you put in the pot for the lung -- let's call 10:57:34
them the lung cancer persons, can I? 10:57:36

A. Okay. 10:57:38

Q. Would include dollars expended for heart disease and 10:57:44
stroke and all the other diseases in the CHD/stroke 10:57:48
category, correct? 10:57:48

A. So long as the person had lung cancer or COPD, so 10:57:56
long as a person had lung cancer or COPD. All of 10:58:00
that person's expenditures would go into the lung 10:58:06
cancer/COPD total. 10:58:06

Q. So that could be heart disease. It could be 10:58:08
stroke. It could be laryngeal cancer. It could be 10:58:12
the diseases that fall under the heading diminished 10:58:16
health status, as well? 10:58:18

A. Correct. 10:58:18

Q. And it could be car accidents, broken bones, and 10:58:24
poisoning, too? 10:58:24

A. Correct. There may -- Dr. Wyant is the person who 10:58:30
has worked with the billing data. I know there were 10:58:32
some exclusions, but I don't know specifically. So 10:58:36
subject to any exclusions that were done, correct. 10:58:40

Q. Okay. Now, an alternative way to approach this 10:58:50
would have been to limit the pot to expenditures for 10:59:00
lung cancer/COPD. 10:59:06

I'm not asking you if that was feasible, 10:59:08
I'm just saying in the abstract that is an approach 10:59:10
which I'm now going to ask you about, okay? 10:59:14

A. (No response.) 10:59:16
Q. I understand why that's difficult for you. Let me 10:59:20
just ask it this way: Why didn't you simply limit 10:59:22
the pot to dollars spent on lung cancer/COPD? 10:59:26
A. Again, Dr. Wyant would be the person who was most 10:59:30
knowledgeable about the billing data. But so far as 10:59:34
I understand it, in my conversation with him, it's 10:59:38
not possible to disaggregate all of the bills and 10:59:42

assign them to a particular condition. 10:59:46
And you'd have to really speak to him 10:59:48
further about, you know, why that's not possible or 10:59:52
what the issue is. 10:59:52
But when you take all of the bills for a 10:59:54
particular person who is suffering from lung cancer 10:59:58
or COPD, it's not possible to say this bill is for 11:00:02
their lung cancer but that bill over there is for 11:00:04
something else. 11:00:06
Sometimes I guess there's information 11:00:06
about that, but it's not sufficient to actually make 11:00:10
the split that you're suggesting. 11:00:12
Q. Now, have you done an analysis, or has someone done 11:00:46
it, of the number of dollars that go into the lung 11:00:56
cancer/COPD pot that are not -- that are clearly not 11:00:58
lung cancer/COPD dollars? 11:01:02
A. I've not done that analysis, no. 11:01:04
Q. Well -- 11:01:10
A. Can I just clarify one thing? I want to clarify the 11:01:12

word "pot" again, I'm sorry. The pot is the thing 11:01:16
before any reductions get applied; is that correct? 11:01:18

Q. Yes. 11:01:20

A. No, I've not done that. 11:01:22

Q. It's funny that you want to use such a silly term, 11:01:26
but, okay, I'll go with it. 11:01:28

A. Tricked again. 11:01:30

Q. Good-natured answer, thank you. 11:01:36
The reports of the defense experts in this 11:01:52
case, which I know you've looked at briefly, 11:01:56
suggests that generally something like 50 percent of 11:02:02
the dollars in the pot, whether it's the lung 11:02:10
cancer/COPD pot or the heart disease/stroke pot, are 11:02:12
not lung cancer/COPD or heart disease/stroke 11:02:16
dollars. Did you see that report? 11:02:18

A. I did not, no. 11:02:18

Q. Let's assume for a moment that that is accurate or 11:02:32
that it's reasonably accurate. Would that be a 11:02:34
matter of concern for this computation? 11:02:36

A. First, I have no, you know, basis for assuming that 11:02:42
that is accurate. 11:02:42

Q. I understand. 11:02:44

A. But I have no reason to be concerned whether it was 11:02:48
or not. 11:02:48

Q. Okay. Now, to go back to the subject we were 11:03:08
talking about a moment ago, which is the third 11:03:14
reduction step. 11:03:28
Let me ask you this question: Suppose 11:03:38

that the dollars in the pot were just lung 11:03:46
cancer/COPD dollars. You with me? 11:03:48

A. This is a hypothetical? 11:03:52

Q. It is a hypothetical. I'm not arguing that it was 11:03:54
doable, I'm just saying suppose they were. 11:03:58
Would the third step, third reduction 11:04:00
step, have still been necessary? 11:04:02

A. Well, let me say in that -- 11:04:20

Q. Take your time. 11:04:30

A. Just repeat it for me one more time. I'm sorry, I'm 11:04:34
trying to understand. 11:04:34

Q. I tried to work up to it so that we had got 11:04:38
everything on the table. 11:04:38
Let's assume that the pot is just 11:04:42
expenditures on the disease under study, lung 11:04:50
cancer/COPD, okay? You with me on the premise? 11:04:54

A. I disagree with the premise. 11:04:56

Q. I understand. 11:04:56

A. And the only reason I disagree is because the 11:05:00
premise is not possible in this context. So, yeah, 11:05:06
that's, anyway -- 11:05:08

Q. Well, you know, the people who have been running the 11:05:12
SAMMEC for all these years are going to be kind of 11:05:16
surprised because this is what they do, they take 11:05:18
their smoking attributable percentage for lung 11:05:20
cancer and they apply it times dollars spent on lung 11:05:24
cancer. 11:05:24

But let's just assume -- I'm not -- don't 11:05:28

want to argue with you about whether it's possible 11:05:32
or not. I want to know why the third step was 11:05:34
necessary. That's the point of my question. 11:05:36

So think of it as an intellectual 11:05:42
exercise, and the point is the third step. It is 11:05:46
not whether or not you can collect the dollars in 11:05:48
the pot differently. Okay? Are you at least 11:05:54
hearing me? 11:05:54

A. I hear you. 11:05:56

Q. That's good. Very small comfort. 11:05:58

A. I'm worried if I don't communicate that I hear you, 11:06:04
but I hear you. 11:06:04

Q. My question is this: Had the pot been just lung 11:06:10
cancer dollars, would the third step reduction still 11:06:14
be necessary? 11:06:16

MR. HAMLIN: Objection to foundation. 11:06:18

THE WITNESS: It's not possible to -- in 11:06:28
the problem that we worked on, it's not possible to 11:06:34
create a pool that has only lung cancer dollars, 11:06:38
which is what made the third reduction necessary. 11:06:44

BY MR. SILFEN:

Q. Okay. Let me tell you what was on my mind. I could 11:07:06
think of two reasons for doing the third step. One 11:07:10
is that we've put dollars into the pot that are not 11:07:20
disease specific and now we have to take some of 11:07:22

them out. That's one reason. 11:07:26
And I take it that you have just affirmed 11:07:28
that that is the reason for the third step. I'm not 11:07:34
going to ask you to say yes because it doesn't 11:07:36
really matter. 11:07:36
I could think of another reason for the 11:07:40
third step, okay. Suppose the purpose of the 11:07:44
calculation were to put the state in the position it 11:07:50
would have been had people not smoked or smoked 11:07:56
differently. Okay? Are you hearing me, as we say? 11:08:00
A. Yep. 11:08:02
Q. Then I can see the third step also, because what you 11:08:12
would be saying is if these people didn't get lung 11:08:16
cancer, they would still have had some medical 11:08:22
condition and it would have cost the state money and 11:08:24
we have to take that into account. 11:08:26
Do you understand what I'm saying? 11:08:28
A. I don't know if you've asked a question. I hear 11:08:32
what you're saying. Is there a specific question? 11:08:34
Q. Was that latter analysis that I just went through 11:08:36
part of your thinking in doing the third step? 11:08:40
A. Would you repeat the latter analysis again? 11:08:42
Q. No. Maybe we'll come back to it. Let's just affirm 11:09:22
a couple of things that I think will probably be 11:09:24
easy. 11:09:26

In the core model, the first two reduction 11:09:28
steps are done based on NMES, correct? 11:09:40

A. The probabilities used in the first two steps are 11:09:42
estimated from the NMES data, yes. 11:09:46

Q. And as we said before, as you said before, the 11:09:56
factors controlled for or taken into account are 11:09:58
age, gender, insurance status, and disease? 11:10:00

A. Correct. And also I think based also whether 11:10:14
someone was in -- there's sort of an adjustment for 11:10:18
nursing home, as well. 11:10:26

Q. I don't know what that is, but -- 11:10:28

A. It's in a footnote, I think. 11:10:32

Q. Now, when you say you controlled for insurance 11:10:58
status, what do you mean by that? How was that 11:11:02
done? 11:11:02

A. To the best of my recollection, we distinguished 11:11:06
people who received their health care from their 11:11:18
employer or their union from those who were on 11:11:20
public assistance in the application. 11:11:22

Q. Why did you do that? 11:11:28

A. We were attempting to apply the core model to three 11:11:36
pots, the Medicaid pot, the GAMC pot, and the Blue 11:11:46
Cross/Blue Shield pot, and so we made our 11:11:50
calculations specific to those pots. 11:11:52

Q. And that is, as you stated earlier, I take it, 11:12:00

because you thought that the results for the 11:12:06
relevant populations might be different? 11:12:08

A. We had the information and tried to make a 11:12:18

calculation specific to these groups as defined by 11:12:20
the variables I indicated, yes. 11:12:22

Q. Well, if that's so, why did you not examine these 11:12:34
separate populations in the refined model, the 11:12:42
diminished health status model, and the nursing home 11:12:46
model? 11:12:46

A. I'm sorry, I don't understand your question. Why 11:12:50
did we not examine these populations? 11:12:54

Q. These separate populations. Why didn't you control 11:12:56
for insurance status? 11:12:58

A. In some applications we did control for insurance 11:13:00
status. 11:13:00

Q. What application? 11:13:02

A. I'd have to look at the specific equations. I don't 11:13:06
recall exactly. 11:13:08

Q. You don't recall any, can't recall any? 11:13:14

A. I recall the insurance status was a variable in some 11:13:18
of the regressions. I don't recall which of them it 11:13:22
was a variable. 11:13:22

Q. Well, it was a variable, but you did not limit the 11:13:26
analysis to that population? In the core model, you 11:13:32
limited your analysis to the public aid population, 11:13:34

correct? 11:13:36

A. No, we did an analysis for Medicaid, GAMC, and for 11:13:44
Blue Cross/Blue Shield. We did not limit our 11:13:48
analysis to the public. 11:13:48

Q. I have a feeling that you know what I meant. When 11:13:50

you were looking for the smoking attributable 11:13:52
fraction for the Medicaid or public aid population, 11:13:56
you did your analysis of just the public aid 11:13:58
population, correct? 11:14:00

A. Yes. 11:14:00

Q. Why didn't you -- 11:14:02

A. To the best of my recollection. 11:14:04

Q. Why didn't you do that in the refined model, the 11:14:08
diminished health status model, or the nursing home 11:14:10
model? 11:14:10

A. And my answer, which I think I gave previously, was 11:14:18
that to the best of my knowledge we did account for 11:14:20
insurance in those -- in some of those models that 11:14:24
you have just referred to. 11:14:26
I don't know if we've referred to -- if we 11:14:30
included in all of the models, I'd have to look back 11:14:32
at the results. 11:14:32

Q. Do you consider that you have accounted for 11:14:38
insurance status when you have it as a variable in a 11:14:40
regression? 11:14:42

A. You are accounting for it, yes. 11:14:44

Q. Are you accounting for it in the same way that you 11:14:48
are when you do your entire analysis in the Medicaid 11:14:54
population, for instance? 11:14:56

A. I don't think I understand your question. 11:15:04

Q. Well, it's this: If you -- you take what you did in 11:15:10
the core model? 11:15:12

A. Right. 11:15:12

Q. Where you did your analysis of conditional 11:15:16
probabilities, if you want to call it that, on a 11:15:20
Medicaid population. Okay? 11:15:22

A. Okay. 11:15:22

Q. And let's suppose that in the refined model you do a 11:15:28
regression on the whole NMES population, all 24,000 11:15:36
people, and you have insurance as a covariate. Do 11:15:40
you consider those to be the same thing? 11:15:42

A. No, but the core model was never intended to be the 11:15:46
same as the refined model. They're not the same 11:15:48
thing. 11:15:48

Q. Just to make sure, I think you answered this 11:16:22
question, but let me make sure. Let's suppose that 11:16:24
you did a regression on just a Medicaid population, 11:16:32
okay? 11:16:32

A. Yes. 11:16:34

Q. And then you did a regression on the whole NMES 11:16:40

population, all 24,000, and had an insurance 11:16:44
covariate. Do you consider those to be the same? 11:16:46

A. No, they're not the same. One is a regression on a 11:16:50
subset of the people in the other. 11:16:52

Q. I took that to be your answer, but he didn't think 11:16:56
so. Forced us to do a second round. 11:17:00

THE WITNESS: He's a statistician, though, 11:17:02
right? 11:17:02

MR. SILFEN: Oh, oh. Does that mean he 11:17:08
was right? 11:17:08

THE WITNESS: I don't know.

BY MR. SILFEN:

Q. Well, actually, it then comes back to the same 11:17:16
question. I mean, your answer was, well, they're 11:17:18
not the same model. 11:17:18

But if the Medicaid population, for 11:17:22
instance, was the proper population in which to 11:17:24
figure the Medicaid smoking attributable fraction, 11:17:28
why didn't you do the refined model on the Medicaid 11:17:30
population? 11:17:32

A. I wouldn't agree that, you know, I mean, you 11:17:36
premised it by saying it was the right analysis to 11:17:40
do it. I would not agree with that. 11:17:42

One major difference between the refined 11:17:44
model and the core model is that the refined model 11:17:46

includes many other variables by which the Medicaid 11:17:50
population may well be different than the Blue 11:17:52
Cross/Blue Shield population. 11:17:54

For example, measures of socioeconomic 11:17:56
status and the like. 11:18:04

Q. Your first answer to me, though, is they weren't 11:18:08
intended to be the same model. What was your 11:18:14
intention in doing a core model? 11:18:16

A. The intention in creating the core model was to make 11:18:24
a simple and more easily understood calculation that 11:18:34
could be compared with the results from the more 11:18:40
complex refined model as a validity check, so we 11:18:46
would feel more comfortable that the refined model 11:18:48

wasn't missing something or didn't need additional work. 11:18:52 11:18:54

And to make clear for ourselves the components of the refined model, that is, the three reductions inherent in that model. 11:18:58 11:19:04 11:19:08

Q. If you want to know the smoking attributable fraction for a Medicaid group, wouldn't it always be preferable to take your measurement in a Medicaid group? 11:19:32 11:19:40 11:19:44 11:19:44

A. What measurement are you referring to? 11:19:48

Q. The measurement of the relative incidence of expense between smokers and nonsmokers? 11:19:54 11:19:56

A. No. 11:19:56

Q. It wouldn't be preferable? 11:20:00

A. You asked would it always be preferable, and the answer is no. 11:20:04 11:20:06

Q. Would it be preferable in this case? 11:20:08

A. Not necessarily. 11:20:10

Q. Well, under what circumstances would it not be preferable to measure the relative health care utilization of smokers and nonsmokers for a Medicaid population in a Medicaid population? 11:20:16 11:20:20 11:20:24 11:20:26

A. When there was limited information about the Medicaid population and useful relevant information in an otherwise similar population from which you can borrow strength. 11:20:32 11:20:36 11:20:38 11:20:40

Q. And where is it that you think there was limited 11:20:42

information about the Medicaid population? 11:20:44

A. You were asking me, I thought, a hypothetical. 11:20:48

Q. Well, I was. But does the hypothetical have any 11:20:52
application here? Is it your view that you had 11:20:54
limited information on the Medicaid population? 11:20:56

A. I used the word limited to mean less than is 11:21:00
available if you use a bigger population. 11:21:04
Yes, that is the case here. We have more 11:21:06
information available than just for those people who 11:21:08
were on Medicaid. We have information on some 11:21:10

30,000 people. 11:21:12

Q. So it's not information, it's the number of people, 11:21:16
sample size you're talking about? 11:21:18

A. Well, sample size as applied to, you know, it's more 11:21:24
than just sample size. 11:21:24

Q. Well, but are you suggesting that you had more 11:21:30
substantive information available for persons not on 11:21:36
Medicaid than for persons on Medicaid? 11:21:38

A. No, I wasn't suggesting that. 11:21:40

Q. Then what are you saying other than sample size? 11:21:46

A. What I'm saying is that in statistical estimation, 11:21:52
it is often better to fit a model that uses more 11:22:00
information on a broader group of people in order to 11:22:06
make an estimate for a particular group. 11:22:10

Q. I don't know what you mean by more information. Do 11:22:12
you mean more quantity or different information? 11:22:14

A. Well, for example, if I wanted to know the 11:22:20
probability of having lung cancer among smokers and 11:22:24

I had two smokers in this room, it would be unwise for me to try -- to ask them whether they had lung cancer, and then to make my estimate for the State of Minnesota, for example, from people from this room. 11:22:30 11:22:34 11:22:38 11:22:40 11:22:42

Whereas, if I had additional information, say from NMES, even for the entire country, that 11:22:44 11:22:48

might be a better way to go about estimating the rate of -- 11:22:52 11:22:52

Q. Am I hearing something other than the number of people, the sample size? I think I'm hearing the sample size? 11:22:54 11:22:58 11:22:58

A. In the context of estimating a simple rate, you can interpret what I said strictly in terms of sample size. 11:23:02 11:23:04 11:23:04

In the context of a regression analysis, the amount of additional information that comes from having more people depends -- is different depending on what their covariate values are, is different from what the predictor variables are. 11:23:08 11:23:10 11:23:12 11:23:16 11:23:20

Certainly, I'm referring largely to sample size. 11:23:22 11:23:22

Q. I want to get something absolutely clear, that your calculations in your regressions are based on values from NMES, right? 11:23:24 11:23:34 11:23:36

A. Correct. 11:23:36

Q. And the values from NMES are the same for the 11:23:40

Medicaid subpopulation and for the rest of the 11:23:44
population, right? 11:23:44

A. You say -- 11:23:46

Q. The information available. 11:23:50

A. Well, the information available, the variables which 11:23:54

are available, may be the same. But the values that 11:23:56
those variables take, for example, the age 11:24:00
distribution, or the gender distribution, may be 11:24:04
quite different for one population versus another. 11:24:06

Q. Okay. Well, for reasons other than sample size, I 11:24:12
mean, if they're different because it's a Medicaid 11:24:16
subpopulation you want it to be different, don't 11:24:18
you? 11:24:18

Isn't that exactly why you want to do it 11:24:20
on a Medicaid subpopulation? 11:24:22

A. I don't want it to be one way or the other. I don't 11:24:24
have a -- I mean, I have no expectation about that. 11:24:28

Q. Were you a part of discussions about whether or not 11:24:34
the refined model and the diminished health status 11:24:40
model and the nursing home model should be done in 11:24:46
insurance subgroups? 11:24:46

A. The models would be the -- I don't recall specific 11:24:56
discussions. 11:24:58

Q. Okay. Now let's talk about the refined model. When 11:26:14
I was working with Dr. Miller on the refined model, 11:26:18
it was sometimes useful to use a paper that he had 11:26:20
prepared. It was marked as Defendants' Exhibit 2324 11:26:26
at the Miller deposition. 11:26:30

Do you want to take a look at that and 11:26:32
tell me if you recognize it or if perhaps you even 11:26:34

reviewed it when you looked at the Miller 11:26:36
deposition? 11:26:38
A. I did not review it when I looked at his 11:26:40
deposition. 11:26:42
Q. Had you seen it before then? 11:26:44
A. I've seen a document that has some of this in it. I 11:26:50
don't know if it's this one in particular. 11:26:52
Q. The important question would be this: On page 13 of 11:27:02
this document, which is Leonard Miller 2324, we see 11:27:12
what was identified by Dr. Miller as equation 3.1. 11:27:18
Do you see that in the middle of the 11:27:20
page? 11:27:20
A. I do. 11:27:22
Q. And he identified that as the -- the major portion 11:27:30
of the ultimate SAF formula in the refined disease 11:27:38
model. 11:27:40
Do you recognize it as that? Or does it 11:27:44
look to you like it looks to me, a mess? 11:27:46
A. It looks like the correct calculation to make to 11:27:52
calculate the smoking attributable fraction. 11:27:56
Q. Good, because that will allow us to work through 11:28:04
that. 11:28:04
Let me ask you to go back again to the 11:28:06
pot, though. We agreed that the -- well, by the 11:28:24
way, let me start this way: The pot total for the 11:28:28

refined model is calculated the same way as it was in the core model, correct? 11:28:34 11:28:34

A. For the specific diseases? 11:28:38

Q. Yes. For lung cancer and COPD, you identified people who had those conditions and then for them you took all their costs incurred in a year? 11:28:42 11:28:44 11:28:48

A. Correct. 11:28:48

Q. And we agreed previously that for a lung cancer person that that could include costs for CHD, if he had it during that year, for other cancers, for diminished health status, and maybe even for car accidents. 11:28:58 11:29:02 11:29:04 11:29:10 11:29:12

That's the conversation we had previously, correct? 11:29:14 11:29:14

A. And for going to the doctor to -- 11:29:18

Q. For a hangnail? 11:29:18

A. For anything, yeah. 11:29:20

Q. Now, when we come to the third reduction step in the refined model, what we're going to subtract is the expected expense for a nonsmoker without a major smoking-related disease, correct? 11:29:26 11:29:32 11:29:40 11:29:42

A. Are you referring to a particular place in the report or into this equation? 11:29:46 11:29:48

Q. I'm referring to the principle of the third reduction. 11:29:50 11:29:52

A. Actually, the second and third reductions are 11:30:04
combined in the refined model equation 3.1, in this 11:30:12
document you've handed me, splits the SAF into two 11:30:16
pieces, the first piece which corresponds to the 11:30:18
first reduction as was done in the core model. 11:30:22

And then a second piece that corresponds 11:30:26
to the second and third reductions taken together. 11:30:30
I wouldn't be able -- I don't think I could 11:30:32
disaggregate that third reduction based upon the 11:30:34
simple quantity -- 11:30:36

Q. You have laid down a weighty challenge here. Now 11:30:40
we're going to go through this equation. 11:30:42

A. Great. 11:30:42

Q. Here we go. Okay. You've agreed that equation 3.1 11:31:22
appears to be the correct formula. Can you 11:31:26
translate it for me into narrative English? 11:31:34

A. Yes. 11:31:34

Q. Okay. Let's do it. 11:31:36

A. Would you like me to narrate it? 11:31:40

Q. Yes, I would like you to do it. 11:31:42

A. Well, actually equation 3.1 is the expected smoking 11:31:50
attributable expenditures, not the smoking 11:31:54
attributable fraction. 11:31:56

Q. It lacks the denominator? 11:31:58

A. Right. 11:31:58

Q. We'll put that aside for the moment. 11:32:00

A. Okay. Well, it has two pieces. The first piece is the rate of smoking among people with the disease. Here written, Prob, open parenthesis, [smokei=1, vertical line, open curly bracket {currtrdi=1}], unclose curly bracket, close bracket. 11:32:06 11:32:16 11:32:24 11:32:26

That first piece before the star represents in narrative English the fraction of people among the fraction of people with the disease who are smokers. 11:32:32 11:32:36 11:32:40 11:32:42

Q. Okay. That's the first reduction step? 11:32:46

A. That is the first reduction step as laid out in the core model. 11:32:48 11:32:48

Q. All right. 11:32:52

A. Okay. And then there's a second piece to the equation, which starts with the third line and goes to the end. 11:32:56 11:33:00 11:33:00

Q. Right. 11:33:04

A. Okay. And that piece compares the expenditures among people who are -- among smokers who are currently treated. That's the first term on the -- the term on the third line. 11:33:10 11:33:14 11:33:20 11:33:22

Q. The expense for a smoker with lung cancer or with a smoking-related disease? 11:33:28 11:33:28

A. Right. It compares that expense to the expense that 11:33:32

would be expected if there were not -- in a population of nonsmokers. 11:33:36 11:33:38

And that expected expense in the population of nonsmokers has two pieces because in 11:33:40 11:33:44

the population of nonsmokers we may have some people 11:33:48
who have lung cancer. 11:33:50

So we take their expenditures and we 11:33:54
average them with the expenditures for people who 11:33:58
aren't -- who don't have lung cancer. 11:34:00

Q. Tell you what, I was trying to catch up. Can you 11:34:02
just go back and -- I think we agreed that we start 11:34:10
with the expense for a smoker with lung cancer. 11:34:14

And you were going to -- could you go from 11:34:16
there? I really got lost, I apologize. 11:34:18

A. Yes. You take the average expense for the group of 11:34:22
smokers with lung cancer and we subtract away a 11:34:28
quantity. So that quantity is the average 11:34:34
expenditures for the population of nonsmokers. 11:34:44

And that average expenditure is calculated 11:34:50
from two kinds of people, a fraction of them 11:34:54
actually have lung cancer, and they have -- tend to 11:34:58
have different expenditures than the other group 11:35:00
which don't have lung cancer. 11:35:02

And this thing we're subtracting away is a 11:35:06
weighted average of those two sets of expenditures 11:35:08

weighted by the fraction of people who actually have 11:35:12
lung cancer in the nonsmokers, appropriately 11:35:14
weighted. 11:35:16

Q. That was excellent. I actually understand it more 11:35:24
clearly than I did before. 11:35:42

We have the first term, which we agreed is 11:35:46

the first reduction, right? 11:35:46
A. The one on the second line, the probability of smoke 11:35:50
given current treated? 11:35:50
Q. Yes.
A. Yes, that's the same as the so-called first 11:35:54
reduction in the core calculation. 11:35:56
Q. And then we're going to multiply that first 11:35:58
reduction times a quantity? 11:36:00
A. Correct. 11:36:00
Q. And the quantity -- I know I'm repeating in part 11:36:04
what you said, but I -- as my wife will tell you, I 11:36:08
don't understand anything until I say it. 11:36:10
The quantity begins with the expense for a 11:36:18
smoker with the condition, a smoking-related 11:36:24
disease?
A. Correct. 11:36:24
Q. And now we're going to make two subtractions? 11:36:28
A. You could think of it as making two subtractions. 11:36:32
Q. The reason why I'm going to try and think about it 11:36:34

that way is I'm going to try and make these fit into 11:36:36
the second and third reductions. 11:36:38
A. Okay. You can do that. 11:36:42
Q. I may learn that I can't, but that's what I was 11:36:48
going to try and do. The first reduction has an 11:36:56
expense term and a weighting or probability term, 11:37:04
correct? 11:37:04
A. Correct. 11:37:04
Q. And the expense term is the average expense for a 11:37:08

nonsmoker with lung cancer? 11:37:10

A. Correct. 11:37:12

Q. And the weighting term is the likelihood of a 11:37:36
nonsmoker getting cancer, is that correct, or 11:37:40
getting a smoking-related disease? 11:37:40

A. It's the fraction of -- it represents the fraction 11:37:44
of nonsmokers that have the disease, correct. 11:37:48

Q. And that is the same as the lung cancers that you 11:38:04
would expect the smokers to get if they didn't 11:38:10
smoke, right? 11:38:12

A. I'm sorry, say that again. 11:38:16

Q. We've identified here the baseline rate of expected 11:38:20
lung cancer among nonsmokers? 11:38:24

A. Yes, this probability represents the fraction of 11:38:30
lung cancers which you tend to get among a 11:38:34
population of nonsmokers, right. 11:38:36

Q. And that is the same as the proportion of 11:38:58
smoking-related disease that we would expect among 11:39:04
smokers even if they didn't smoke, correct? 11:39:06

A. We've had -- 11:39:10

Q. Our semantic problem. 11:39:12

A. We've had this discussion. 11:39:12

Q. We have a semantic problem, but semantics aside 11:39:16
you'd agree with me? 11:39:16

A. I would answer it the same way I answered it 11:39:20
before. 11:39:22

Q. All right. Isn't this, in a sense, the how much 11:39:34

http://legacy.library.ucsf.edu/tid/gqq07a00/pdf

extra disease reduction? 11:39:36
A. No. 11:39:40
Q. Why not? 11:39:44
A. In the core model, the how much extra disease 11:39:52
reduction is the difference between the disease rate 11:40:02
for the smoking population minus the disease rate 11:40:04
for the nonsmoking population divided by the disease 11:40:10
rate for smoking population. 11:40:10
This is one of the terms in that 11:40:12
reduction, but it's not the same as that reduction. 11:40:20
Particularly, it's multiplied by an expenditure, as 11:40:24
well. 11:40:24
Q. The third term is the expense for a nonsmoker 11:40:54
without the condition, correct? 11:40:56

A. The third -- let's see the -- the last term -- 11:41:04
Q. Second subtraction, I'm sorry. 11:41:06
A. Yes, that has expenditures for persons who are 11:41:10
neither smokers nor currently treated for the 11:41:12
smoking attributable disease. 11:41:12
Q. Well, as you've just described it, why isn't that 11:41:18
the third reduction? 11:41:18
A. Well, that's a component of the third reduction. 11:41:26
The other component is up in that second piece we 11:41:32
just talked about. 11:41:32
The other component of the third reduction 11:41:44
is the expenditures for people who -- I'm sorry, the 11:41:44
expenditures for people who have the disease -- 11:41:52
let's see, you've got me confused here. 11:41:54

http://legacy.library.ucsf.edu/tid/iqq07a00/pdf

Q. I'm glad you're confused. That's very comforting. 11:42:00
A. Yeah. 11:42:00
Q. Why don't we leave it until after the break. I'd 11:42:12
like to think about what you said and maybe it will 11:42:14
sink in and I'll see it. 11:42:16
It is clear, however, is it not, that in 11:42:42
these two subtraction steps, we are accomplishing 11:42:48
the reductions described as your second and third 11:42:58
reductions? 11:43:00
A. Yes. 11:43:02
Q. Good. Now, let's go back to the point we were 11:43:06

working on before. If we go back to your 11:43:20
description of the third reduction on page 8, the 11:43:50
quantity we're identifying there is the average 11:43:54
expenditures of never-smokers without the disease, 11:43:58
correct? 11:43:58
A. We're on page -- 11:44:00
Q. That's at the very end of the how many more 11:44:06
dollars. We're going to turn that into a 11:44:08
proportion? 11:44:10
A. We take the difference between the average 11:44:12
expenditures for smokers with the disease, which is 11:44:16
the first term here. 11:44:18
Q. Yes. 11:44:20
A. And we subtract away from that the average 11:44:22
expenditures of never-smokers. 11:44:24
Q. Without -- 11:44:26

A. Without the disease. 11:44:28
Q. Okay. Now, we agreed previously that when we were 11:44:32
putting our pot together for a person with lung 11:44:40
cancer, forgive me for repeating this, but the 11:44:46
record probably wouldn't read if I didn't, for the 11:44:48
person with lung cancer, that person in that year 11:44:50
may also have CHD, may also have other cancers, may 11:44:54
also have diminished health status, may also have 11:44:58
other costs, car accidents, doctor visits, right, 11:45:00

and they all go into the pot? 11:45:02
A. Correct. 11:45:04
Q. Now, when we identify expenditures, the average 11:45:10
expenditures expected for a nonsmoker without a 11:45:16
major disease -- 11:45:16
A. Right. 11:45:16
Q. -- isn't it correct that we are going to leave in 11:45:24
that pot the lung cancer/COPD costs, the CHD/stroke 11:45:32
costs, all of the other major smoking-related 11:45:36
disease costs, and any diminished health status due 11:45:40
to smoking? 11:45:40
A. That's my understanding, yes. 11:45:48
Q. Now -- 11:45:58
A. And that due to smoking modified all the things that 11:46:02
were left in. 11:46:02
Q. Yes, I agree. Well, let me back up on that. In a 11:46:18
way I think that puts the question I'm asking. 11:46:24
To get to the smoking attributable portion 11:46:28
for a lung cancer person, you did an analysis for 11:46:36

your first and second reductions of the relative 11:46:40
incidence of lung cancer among smokers and 11:46:44
nonsmokers, correct? 11:46:44

A. The relative, the rate of lung cancer among smokers 11:46:52
and nonsmokers, as well as the fraction of smokers 11:46:56
among the lung cancer -- among people who had lung 11:47:02

cancer. 11:47:02

Q. So those reduction steps which get you to a smoking 11:47:06
attributable amount are based -- 11:47:08

A. First two steps. 11:47:10

Q. -- are based on lung cancer? 11:47:12

A. Are based upon persons who have, among other things 11:47:16
that they might have, lung cancer and COPD. 11:47:20

Q. I know, but the percentage is based on the number of 11:47:24
smokers among the people with lung cancer? 11:47:28

A. Who have at least lung cancer? 11:47:30

Q. Yes. 11:47:30

A. And COPD, right, or COPD, and may have other 11:47:34
things? 11:47:34

Q. Right. 11:47:34

A. Right. 11:47:36

Q. Well, then, doesn't that mean that you are going to 11:47:38
be applying first and second reduction steps that 11:47:42
are based on lung cancer to dollars that are spent 11:47:48
on heart disease and diminished health status? 11:47:52

A. That's a possibility. 11:47:54

Q. Is that something that you fellows discussed? 11:48:00

A. We have discussed it. 11:48:02

Q. And I may be running into a work product objection 11:48:10
here, but is -- what have you discussed? 11:48:12

A. Well, we discussed what the options were. The 11:48:22

options were laid out and we tried to identify the 11:48:26
best strategy, the fairest strategy, given the 11:48:30
options, and this is what we did. 11:48:32

Q. I take it you're talking about conversations that 11:48:36
occurred before the report was put in? 11:48:38

A. Some of them occurred before and some of them 11:48:40
occurred subsequently. 11:48:42

Q. What were the options that were discussed? 11:48:44

A. Well, we were unable to separate the expenditures 11:48:52
for a person, for example, who has both lung cancer 11:48:54
and suffered a stroke. 11:48:58

It's hard for us to separate out how many 11:49:00
dollars go to the lung cancer and to the stroke, so 11:49:04
far as I understand from Dr. Wyant. 11:49:08

So we could have taken all of those, and 11:49:10
I'm using that as an example. I don't know that 11:49:12
particular case. 11:49:14

Q. I understand. 11:49:14

A. Let me, in fact, correct myself. We have a person 11:49:18
who has lung cancer and some other smoking 11:49:20
attributable disease. It's difficult to sort out 11:49:24
there expenditures as this one is due to that and 11:49:30
this one is due to that. So we had to take them as 11:49:32
a lump. 11:49:36

So we could either allocate those 11:49:38

expenditures to the lung cancer case or we could 11:49:40 allocate them to the other pot. Those were our two 11:49:46 choices. 11:49:46

And, of course, whatever we did we would 11:49:48 do the same thing to be fair, do the same thing with 11:49:50 NMES calculations, as well. 11:49:54

And so this was the choice we made. One 11:50:00 factor -- well, this was the choice we made. 11:50:02

Q. Have you done a sensitivity analysis to -- well, one 11:50:08 I asked you about in the beginning was to see how 11:50:10 many dollars we were talking about. You have not 11:50:14 done that? 11:50:14

A. I have not done that, no. 11:50:16

Q. And since you haven't done that, I take it you also 11:50:18 haven't done an analysis to see what the effect may 11:50:24 be of applying the lung cancer reduction step to 11:50:30 heart disease dollars. You have no idea what that 11:50:32 effect is? 11:50:34

A. I've not done that calculation, no. 11:50:36

Q. Wouldn't it be pretty clear that -- I'm going to do 11:50:52 this A and B. If, A, the nonlung cancer dollars are 11:50:58 sufficiently high, and, B, the difference in the 11:51:02 lung cancer and heart disease and diminished health 11:51:06 status reduction steps is sufficiently high that the 11:51:12 result has been skewed? 11:51:12

A. No, I don't agree. 11:51:16

Q. Understanding that you don't have to accept my A and 11:51:18
B, why wouldn't that be true if my A and B are 11:51:22
right? 11:51:22

A. Well, one reason is that the health care costs for 11:51:28
having disease B alone and the health care costs for 11:51:36
having -- well, let me back up. 11:51:40
The health care costs of having disease A 11:51:42
and B together may be very different than the sum of 11:51:46
the health care costs of having A or having B. 11:51:52
So, for example, if you're trying to treat 11:51:54
a person with heart disease, but that person also 11:51:58
has lung cancer, the costs of their heart disease 11:52:02
may be very much larger. 11:52:06
In fact, Dr. Samet has suggested maybe 11:52:10
very much larger than would be their costs had they 11:52:12
not had lung cancer. And those additional costs are 11:52:14
appropriately attributed to the -- to smoking. 11:52:18

Q. Well, that -- 11:52:20

A. To the extent to which their lung cancer -- 11:52:22

Q. Which they're attributed to the lung cancer? 11:52:24

A. Yes. 11:52:26

Q. You could just as easily reason it the other way, 11:52:28
though, and I know you're not arguing otherwise. 11:52:30

A. Yes. So that's one reason why what we're doing is 11:52:36

valid. Let me be clear. That's one reason why I 11:52:52
wouldn't agree with your premise that you had put in 11:52:56
the question to me. 11:52:56
Q. If you were doing a journal article on this subject, 11:53:10
wouldn't you consider it to be incumbent on you to 11:53:16
measure the impact of this choice that you've made? 11:53:22
A. It wouldn't be incumbent upon me, no. In analysis, 11:53:32
there are many choices that one makes. One does the 11:53:36
best to look at the impact of the various choices 11:53:40
and to do sensitivity analyses. 11:53:44
Q. Perhaps I shouldn't have said incumbent. Wouldn't 11:53:46
it have been expected and probably preferable? 11:53:48
A. Not if -- what I was trying to say in my response 11:53:52
was that you do your best to put in priority the 11:53:56
things that need to be looked at. 11:53:58
And, you know, we are looking at things. 11:54:00
We're continuing to look at things. And we've gone 11:54:04
for the highest priorities, ones based upon our 11:54:06
judgments of the ones that would make the biggest 11:54:10
difference, potentially make the biggest 11:54:12
difference. 11:54:14
Q. Now, in the formula that we've just worked our way 11:55:28
through, I think we have some hangover questions on 11:55:30
it which I'm going to try and think about at lunch. 11:55:36
But where I would go next is where I went 11:55:40

with Dr. Miller, which is to try to understand how 11:55:42
the components of that formula were derived. And 11:55:48

the formula seems to me to be made up of two chunks 11:55:56
of information. 11:55:58

One chunk is the three conditional 11:56:04
probabilities and the other chunk is the three 11:56:10
expense numbers, smoker with, nonsmoker with and 11:56:18
nonsmoker without. 11:56:20

Is that a fair shorthand of what's in 11:56:22
here?

A. You're referring to the piece that has to do with 11:56:26
the second and third reductions? 11:56:28

Q. Actually, I'm referring to everything. 11:56:30

A. Everything, sorry. Yes, there are six terms there 11:56:32
and you've listed them, yes, that's correct. 11:56:34

Q. Well, that's the level of math at which I'm really 11:56:38
good. 11:57:18

Okay. Let's talk about the conditional 11:57:20
probabilities. I'm not sure whether we coined that 11:57:26
term or you did. Is that a term that makes sense in 11:57:28
this context? 11:57:28

A. Conditional probability is a term used in 11:57:34
probability or in statistics. 11:57:38

Q. I know I didn't coin it. 11:57:40

A. Somebody must have. 11:57:40

Q. Somebody must have said it to me. The probabilities 11:57:52
that we're going to derive, let's take lung 11:58:04
cancer/COPD first because in refined model it's done 11:58:06
differently from lung cancer/COPD and CHD/stroke? 11:58:12

A. Correct. 11:58:12

Q. Who made the decision, by the way, to use different 11:58:16
methodologies within the refined model? 11:58:18
A. I would actually call it the same methodology. The 11:58:24
idea of estimating the terms in this equation 3.1. 11:58:30
A different estimation method was used. The 11:58:32
approach or the method, the overall method, is the 11:58:36
same, but the specific approach to estimating these 11:58:38
unknown probabilities was different, yes. 11:58:44
Who made that decision? That was based 11:58:46
upon discussions largely with Dr. Samet. The 11:58:56
decisions are made, you know, by all of us. 11:59:00
Q. Okay. Now, I think of it in terms of three 11:59:08
principal conditional probabilities. Given lung 11:59:12
cancer, what is the probability that you smoke? And 11:59:18
given that you smoke, what is the probability of 11:59:20
lung cancer? And then given that you don't smoke, 11:59:24
what is the probability of lung cancer? 11:59:26
I know those aren't exactly what occur in 11:59:30
the equation, but I think those are the 11:59:32
probabilities that are derived in order to get to 11:59:34

the quantities in the equation. 11:59:36
A. I think it's true that those three conditional 11:59:40
probabilities you stated would be sufficient to get 11:59:44
us to the three that appear in this equation. 11:59:46
Although, I'm not -- 11:59:48
Q. That is a very congenial thing to say, because I 11:59:50
think that's exactly what I meant to say. That I 11:59:54

think those are the three conditional probabilities 11:59:54
that are sufficient to get you to the quantities in 12:00:00
the equation. That is exactly what I meant to say. 12:00:02
But we'll check on that as we go along. 12:00:04
First methodology. For lung cancer, let's 12:00:10
just take one we agree on, given lung cancer what is 12:00:12
the probability that you smoke? That we know you 12:00:16
had to do because that's the first term in the 12:00:18
equation. 12:00:18
A. Correct. 12:00:20
Q. And the way that is done, it's done in NMES, and 12:00:28
it's done for age, gender -- by age, gender, and 12:00:34
it's just a simple count. If there are 1,000 lung 12:00:40
cancers and 900 of them are smokers, it's going to 12:00:44
be 900 over 1,000, or 90 percent? 12:00:48
A. I think the -- I think that's correct. 12:00:54
Q. Okay. And similar for the other conditional 12:00:58
probabilities, it is simply an age/gender count in 12:01:04

those? 12:01:04
A. I'm going to just reserve my answer. I can't 12:01:12
remember exactly what the strata were. I think it's 12:01:18
just age and gender, but I don't have a perfect 12:01:20
recollection of that. 12:01:20
But that's right, it's the frequency with 12:01:22
which these events occur in the population as 12:01:26
indicated by the conditioning in these 12:01:30
probabilities. 12:01:32
Q. And for CHD/stroke, these conditional probabilities 12:02:00

are derived by a form of regression, correct? 12:02:06
A. Correct. 12:02:08
Q. And tell me the name of the form of the regression 12:02:12
because I always say it wrong. 12:02:14
A. Well, it's a probit regression. 12:02:18
Q. Well, isn't it a binumeral variate joint 12:02:24
distribution, maximum likelihood thing, something 12:02:30
like that? 12:02:32
MR. HAMLIN: Objection to form. 12:02:32
BY MR. SILFEN:
Q. I could get it right if I wanted to, there's just 12:02:38
some things I don't care to learn. I'm too old. 12:02:40
A. Would you mind repeating the question? No, it was 12:02:44
a -- the model was what you might call a bivariate 12:02:50
model, meaning that it was a model for currently 12:02:56

treated and smoking as the two outcomes. It was of 12:03:00
the probit form. 12:03:00
Q. So you did a first regression, which is whether you 12:03:04
are a smoker given your other characteristics? 12:03:08
A. No. 12:03:12
Q. I thought that you did two regressions. In one the 12:03:20
outcome is whether you're a smoker, and the other is 12:03:22
whether you're being currently treated? 12:03:24
A. Well, it's two regression equations fit together 12:03:32
because you're estimating the joint distribution of 12:03:40
these two currently treated and smoking variables. 12:03:44
Q. Okay. But the two regressions fit together are, 12:03:48

first, whether you are a smoker as the outcome as a 12:03:56
function of your characteristics, whatever you were 12:04:00
using as your covariates? 12:04:02

A. Correct. 12:04:02

Q. And the second is whether you are currently being 12:04:06
treated for CHD/stroke as a function of your other 12:04:10
covariates, but not smoking? 12:04:14

A. Correct. 12:04:16

Q. And from that, you derive a correlation coefficient 12:04:20
between disease and smoking? 12:04:24

A. Well, just as the regression parameters in those two 12:04:34
regressions are parameters of this analysis, so, 12:04:38
too, is the correlation coefficient another 12:04:40

parameter. 12:04:42
It's another thing that we estimate when 12:04:44
we do the simultaneous bivariate regression. 12:04:48

Q. It's going to work like the coefficients in the -- 12:04:52

A. Think of it as another parameter to be estimated 12:05:02
from the joint information about current treated and 12:05:06
smoking. 12:05:06

Q. Now, how was it decided to do this bivariate 12:05:14
analysis, and why was it decided to do a bivariate 12:05:18
analysis? 12:05:18

A. Well, what we needed to get out of this analysis 12:05:26
were these probabilities, these three probabilities 12:05:28
that you've referred to that are in equation 3.1 in 12:05:32
this document you've given me. 12:05:40

Q. Right.

A. The fourth probability is determined from the 12:05:42
others. So really what we needed to do was to 12:05:46
estimate the probabilities in a little two-by-two 12:05:50
table. 12:05:50
And so the way to do that is to fit a 12:05:56
model that estimates those probabilities and allow 12:06:00
them to be a function of other factors which we 12:06:04
wanted to control for. 12:06:04
And that's what -- this is sort of a 12:06:06
standard method that's used for that purpose, 12:06:14

standard statistical technique. 12:06:16

Q. I guess -- I understand you answered a question. I 12:06:22
had something different in mind. In NMES -- which 12:06:28
is what you're using here, correct? 12:06:32

A. NMES was used to estimate this bivariate model, 12:06:38
yes. 12:06:38

Q. You have smoking information. You have other 12:06:44
covariates. You have incidence of disease. And you 12:06:52
have expenditure for disease. You have all that 12:06:54
information, correct? 12:06:56

A. Correct. 12:06:56

Q. And the outcome you're interested in here is 12:07:06
expenditure for disease, isn't it, ultimately? 12:07:12

A. Well, that's one of the outcomes. 12:07:14

Q. Well, I'm not talking about the methodology you 12:07:18
decided on here, I'm talking about the end of the 12:07:20
day. At the end of the day, you wanted to know the 12:07:22

expenditures -- 12:07:24

A. Which were attributable to smoking. 12:07:26

Q. -- which were attributable to smoking. 12:07:30

Why wouldn't one have specified a model in 12:07:34
which the outcome was expenditure and the predictors 12:07:40
were the covariates that you were interested in? 12:07:42

A. Because that model wouldn't address the question 12:07:46
that we've addressed in this study. 12:07:48

Q. I think I'm puzzled by that. I don't know why you 12:07:56
would get a -- if smoking were a predictor, you 12:08:00
would get a coefficient for smoking, the 12:08:04
contribution of smoking to predicting expenditure, 12:08:08
correct? 12:08:08

A. You would get a regression coefficient for smoking 12:08:12
if it were a variable in an expenditure model. 12:08:16

Q. And once with that, you could surely determine a 12:08:20
smoking attributable fraction, couldn't you? 12:08:24

A. I don't know. 12:08:24

Q. Well, I mean, I can think of lots of easy ways to do 12:08:32
it, and I'm an English major. For instance, you 12:08:38
could -- you estimate the model on your smokers and 12:08:40
nonsmokers, you get coefficients, and then you put 12:08:46
the smokers back in the model but you change only 12:08:52
one thing, you change their smoking coefficient, 12:08:54
make them nonsmoker, see the difference, why not? 12:09:02

A. You know, I don't know. I mean, I don't know how to 12:09:06
answer it because I didn't do that, and I've not 12:09:10
thought carefully about what you're proposing as to 12:09:12

what its problems might be. It's not something I did or thought carefully about the possibility of doing. 12:09:18 12:09:24 12:09:24

Q. Yeah, but when one looks at the schematic of what you've done here, why? Why did you do all this? 12:09:30 12:09:36

I'm quite serious, why? You had the outcome in NMES, which is expense. You had all the covariates. Why did you do all this? 12:09:40 12:09:42 12:09:48

A. Why did we do the refined model? 12:09:52

Q. Why didn't you just build a model to predict the outcome that was of interest here? 12:09:56 12:09:56

A. We did build a model to estimate the smoking attributable fraction of dollars. 12:10:04 12:10:06

Q. Why didn't you do it in the most obvious and direct way? 12:10:10 12:10:10

A. We did it in an obvious and direct way. 12:10:14

Q. Was it discussed that the information was available to you to estimate a simple model in which the outcome was expense and the covariates were disease and any other variables you were interested in, was that discussed? 12:10:24 12:10:28 12:10:34 12:10:38 12:10:40

A. Our analysis is a regression analysis that estimates the quantities necessary to calculate the smoking attributable fraction. And I'm not going to speculate further about what we -- 12:10:46 12:10:48 12:10:58 12:11:00

Q. That wasn't my question. My question was whether there was -- I'm not asking you to speculate on the 12:11:02 12:11:06

value of another model. 12:11:08
I'm asking you -- really I'm asking you 12:11:10
questions about this model. Why this model? Was it 12:11:12

discussed that you could have specified a very 12:11:14
simple direct model in which the outcome was expense 12:11:20
and the covariates, the predictors, were smoking and 12:11:26
the other variables of interest? Was it discussed? 12:11:28

MR. HAMLIN: Objection; asked and 12:11:30
answered. 12:11:32

MR. SILFEN: No, he hasn't answered.

THE WITNESS: I don't recall a discussion 12:11:32
of the specific model that you've proposed, as I 12:11:34
understand it. 12:11:34

BY MR. SILFEN:

Q. Well, can you tell me what is the advantage of your 12:11:44
model over the one that I suggest? 12:11:46

A. I don't know because I don't know specifically what 12:11:50
model you are suggesting. 12:11:52

MR. SILFEN: Let's break for lunch. 12:11:54

MR. HAMLIN: Do you mind going until about 12:11:58
12:30 or not? 12:12:00

MR. SILFEN: What's up? 12:12:02

MR. HAMLIN: Just that we've got lunch 12:12:06
reservations for a little bit later. 12:12:10

MR. SILFEN: Tom, I'm going to do this for 12:12:18
you. My questions will get a little weak, but 12:12:20
you'll go with me, won't you? 12:12:22

THE WITNESS: I'll do my best. 12:12:26

MR. SILFEN: Let's do it. 12:12:42

(Discussion held off the written record.)

MR. HAMLIN: This is off the record. 12:12:46

THE VIDEOGRAPHER: We are temporarily 12:12:50
going off the video record. The time is 12:12 p.m. 12:13:02

(Off the record.) 12:13:44

THE VIDEOGRAPHER: Back on the video 12:14:14
record. The time is now 12:14 p.m. 12:14:16

BY MR. SILFEN:

Q. I had a piece prepared here where I convinced myself 12:14:22
and everyone else that you could take those three 12:14:24
conditional probabilities and turn them into the 12:14:28
terms in the final equation, and I think you 12:14:34
relieved me of that. 12:14:34

You're pretty sure that you can do that? 12:14:36
Those are the sufficient probabilities, the ones we 12:14:40
discussed, correct? 12:14:40

A. I'm not sure because you said them quickly. And I 12:14:44
think so, but I'm not 100 percent sure. If we wrote 12:14:48
them down on a piece of paper, I could tell you for 12:14:50
sure. 12:14:50

Q. Well, why don't I save -- do you want to write them 12:14:58
down? 12:14:58

A. Can I write them as you say them, that would be 12:15:00
helpful. 12:15:00

Q. We won't belabor it, if you're going to be able 12:15:04
to -- given lung cancer, what is the probability 12:15:08
that you smoke? Given smoke, what is the 12:15:14
probability of lung cancer? And given no smoke, 12:15:16
what is the probability of lung cancer? 12:15:22

I'd like to see how you write those. He's 12:15:28
making a box. Let the record reflect he's making a 12:15:30
box. 12:15:32

A. I believe it's true that these three conditional 12:15:38
probabilities would be sufficient for us to 12:15:40
calculate the joint distribution of lung cancer and 12:15:46
smoking and thereby then calculate the other 12:15:48
probabilities listed in this 3.1. 12:15:50

Q. Thank you. Give me back my pad, but take your piece 12:15:58
of paper. 12:16:00

You can keep it. I don't care. It
doesn't look like it would do me any good.

MR. HAMLIN: Do you want to mark that? 12:16:06

MR. SILFEN: No. It took me seven hours 12:16:08
to do a proof of that. See this? I did this 12:16:16
myself. 12:16:16

MR. HAMLIN: Can we mark that, too? 12:16:26

MR. SILFEN: I've got 15 minutes to work
with, we might as well start marking stuff I did.

BY MR. SILFEN:

Q. Okay. We've done the conditional probabilities, now 12:16:32

we want to do what I think of as the expense model. 12:16:36
And there is an expense regression, correct, in 12:16:50
which the outcome is the probability, and outcomes 12:16:54
are probability and level of expense? 12:16:56

A. Correct. 12:16:58

Q. And what we're talking about there is the 12:17:02
probability of any expense? 12:17:06

A. In the first of the two equations that you 12:17:10
mentioned? 12:17:10

Q. Yeah. 12:17:12

A. Correct. 12:17:12

Q. So once again it's not just an expense related to a 12:17:16
smoking-related disease, it's any medical expense? 12:17:20

A. For a person who has that particular disease, yes, 12:17:26
so if we're talking about the -- well, if we're 12:17:30
talking about the CH- -- 12:17:32

Q. I don't think so. Well, then let's get into that. 12:17:36
The outcome is -- I understand the outcome to be 12:17:38
probability of expense, any expense, yes, no? 12:17:44

A. Yes, that's correct. 12:17:46

Q. And now what we want to know is the components of 12:18:00
that regression. And I understand them to be a 12:18:14
group of covariates, things like race and education 12:18:20
that vary somewhat, depending on their 12:18:24

significance? 12:18:24

A. The model, as originally specified, had a set of 12:18:30
variables and there was a reduction of variables 12:18:32

based upon their procedure to get to a smaller 12:18:36
model, yes. 12:18:38

Q. And then there is a disease term, any 12:18:48
smoking-related disease, any major smoking-related 12:18:52
disease, yes, no? 12:18:56

A. What model are we talking about? I'm sorry. 12:19:00

Q. We're talking about the -- 12:19:04

A. CHD/stroke? 12:19:06

Q. No, my understanding is there is just one regression 12:19:10
and it is not disease specific, it's any -- the 12:19:14
variable, the outcome is probability of expense, 12:19:18
yes, no, or level of expense. And the variable used 12:19:22
is any smoking-related disease, yes, no. 12:19:28

A. But there's a separate modeling effort for the 12:19:30
diminished health and for the disease modeling. 12:19:34

Q. Diminished health is separate? 12:19:36

A. Not talking about that. 12:19:38

Q. Not talking about that, talking about the disease 12:19:40
model. 12:19:40

A. So it's for the diseases, and, I'm sorry, ask your 12:19:46
question again. 12:19:46

Q. Well, I'm trying to understand the components of 12:19:50

disease regression. I think we've agreed that the 12:19:54
outcome is probability of expense, yes, no, that 12:19:56
there are covariates in there, if they survive the 12:20:02
significance test of the order of education or race, 12:20:04
demographics. 12:20:06

And then I'm talking about a disease term, 12:20:08

and I am stating that, as we read the code, the 12:20:14
disease term is any smoking-related disease, any 12:20:18
major smoking-related disease, yes, no, that it is 12:20:24
not disease specific? 12:20:24

A. I'm not sure. I'd have to review. 12:20:32

Q. Okay. Well, isn't that a somewhat important point 12:20:38
because as you've already anticipated, if it's, as I 12:20:44
say, any smoking-related disease, that means your 12:20:46
expense equation does not distinguish between 12:20:50
diseases? 12:20:50

A. Well, it distinguishes only to the level of the 12:20:58
major smoking attributable diseases, if that's 12:21:00
true.

Q. And you have no -- 12:21:02

A. I can't remember the specific equation. It's not 12:21:06
there right now. 12:21:06

Q. Well, I understand that, but this strikes me as 12:21:10
something a little different than remembering a 12:21:12
specific equation. This is a concept. 12:21:14

To your knowledge, was your regression 12:21:18
sufficiently sensitive to tell the difference 12:21:22
between expenses for different diseases? 12:21:26

MR. HAMLIN: Objection; asked and 12:21:26
answered. 12:21:28

THE WITNESS: I don't remember. I don't 12:21:30
recall that specifically. 12:21:30

BY MR. SILFEN:

Q. Okay. Now, there is a -- if you have the materials 12:21:42
available to check that, you might check it because, 12:21:46
of course, then the questions would be, well, why? 12:21:48
And I can't ask you those questions when you don't 12:21:52
know. 12:21:56
There is also a smoking term. Do you 12:22:00
remember what the smoking term is in the expense 12:22:04
regression? 12:22:04

A. I believe it's the conditional expectation of the 12:22:16
smoking variable as specified in the bivariate 12:22:22
model. I'm not sure. Given the reported, the 12:22:28
result, but I'd have to go back and look at the 12:22:30
equation, again. 12:22:32

Q. Why don't I tell you what I believe it is and then 12:22:36
you can check because, obviously, then I'm going to 12:22:38
ask you why you made certain choices. 12:22:40
I believe that there is no main effect of 12:22:44

smoking in the expense equation. 12:22:46

A. This is the probability of expenditure, correct? 12:22:50

Q. Yes, or level of expenditure, either one. I don't 12:22:52
mean this -- I mean this to refer more to my use of 12:22:56
the terms than your knowledge of them. 12:22:56
Do you know what I mean when I say main 12:22:58
effect of smoking? 12:23:00

A. I know -- well, I know what a main effect is in a 12:23:04
statistical model. 12:23:06

Q. Okay. Again -- 12:23:06

A. But I don't know what you mean. 12:23:08

Q. I'm trying to make sure that I'm using the terms 12:23:10
correctly, not testing your understanding. I know 12:23:14
you understand it. 12:23:14

What I think of as a main effect would be 12:23:18
smoking alone as opposed to smoking in an 12:23:20
interaction term. I'm drawing that distinction. Is 12:23:26
that sensible? 12:23:26

A. In some contexts, it is sensible. In this 12:23:30
particular context, what I believe is in the model 12:23:34
are what you might more usefully think of as two 12:23:36
main effects. 12:23:38

One main effect for public and one main 12:23:40
effect for private insurance. 12:23:42

Q. That is my understanding. What is in the regression 12:23:46

are terms for pubsmk or privsmk, depending on the 12:23:52
insurance status of the person? 12:23:54

A. Right. 12:23:56

Q. Why do we have -- well, what do you mean by two main 12:24:02
effects? This is an interaction term, is it not, 12:24:06
statistically, pubsmk? 12:24:12

A. Yes. 12:24:12

Q. Well, then, why do you call it two main effects 12:24:16
rather than two interaction terms? What distinction 12:24:20
are you drawing? 12:24:22

A. Because those two terms are meant to reflect the 12:24:26
effect of smoking on expenditures where you get -- 12:24:34
you allow for there to be one size effect for public 12:24:38

expenditures and a different size effect for private 12:24:44
expenditures. 12:24:46

Both of them are the effects of smoking, 12:24:48
but just it allows it to have a different value for 12:24:52
the two subsets. 12:25:00

Q. Why are you doing that? 12:25:00

A. There's the possibility, as we discussed previously, 12:25:04
there was the possibility that the effects of 12:25:06
smoking might be estimated to be different in these 12:25:08
two subgroups. 12:25:08

Q. Yeah, but you've already got a disease variable in 12:25:14
this regression. Take my word for it that the 12:25:18

variable is any smoking-related disease. Miller 12:25:22
told me it was. On the second day he came back and 12:25:24
he affirmed. 12:25:28

That means that the significance of the 12:25:30
smoking coefficient here isn't the contribution of 12:25:34
smoking to the occurrence of disease, it's the 12:25:38
contribution of smoking to the probability of 12:25:40
expense over and above the effect of having a 12:25:44
disease. 12:25:44

Isn't that right? 12:25:46

A. Over and above the effect of controlling for a 12:25:48
variable that's in that model, which may or may not 12:25:54
accurately and perfectly reflect the condition. 12:25:58

Q. But you're agreeing with me that what you're 12:26:04
measuring here is not the effect of smoking on 12:26:04
disease and then on expense, because you've already 12:26:06

got a variable for disease in here? 12:26:10
A. I've answered. 12:26:12
Q. My question is: When you measured the effect of 12:26:16
smoking on disease in your conditional 12:26:22
probabilities, you didn't limit it to a Medicaid 12:26:26
population, right? 12:26:26
A. In the -- let's see, in the bivariate model? 12:26:32
Q. Yeah. 12:26:32
A. I think that's correct. 12:26:38

Q. So -- 12:26:40
A. Although, I would want to refresh myself to be 12:26:42
sure. 12:26:42
Q. Why do you come here in this expense model with the 12:26:46
effect of smoking on disease already determined and 12:26:50
put in an interaction term? 12:26:54
A. I don't understand your question. 12:26:58
Q. Well, you've measured the effect of smoking or the 12:27:02
correlation of smoking with disease already by the 12:27:04
time you get to this probability of expense 12:27:08
equation, correct? 12:27:08
A. We have, prior to fitting that equation, a model 12:27:16
which allows us to estimate the conditional 12:27:18
probabilities that are necessary in equation 3.1. 12:27:26
Q. Right.
A. We do not yet have a model that allows us to 12:27:28
estimate the expenditures. 12:27:30
Q. I understand. But in your expenditure regression, 12:27:32

you have a disease term, which is any smoking, major 12:27:38
smoking-related disease, yes, no. All right, you'll 12:27:42
accept that? 12:27:42

A. I don't have any choice since I don't remember. 12:27:46

Q. Okay. 12:27:48

A. Although, the best answer is I don't remember 12:27:50
exactly the form of that equation right now, but go 12:27:54

ahead. 12:27:54

Q. Assuming that is the form of the equation, what is 12:28:00
the interpretation of the coefficient of the pubsmk 12:28:10
term? 12:28:10

A. It's the additional expenditure for -- additional 12:28:18
public expenditures for persons who are smokers, 12:28:22
conditioned on the reported value that's in the 12:28:24
equation for currently treated disease. 12:28:28

Q. My question is: Why would you specify your model so 12:28:34
that that value is measured in a Medicaid 12:28:46
population, but the relationship of smoking to 12:28:50
disease is not? 12:28:52

A. Well, it's -- one reason is because it's likely that 12:29:00
expenditures will differ in these two populations 12:29:06
having to do with demand or access or other sorts of 12:29:10
things. Whereas, the -- well -- and that's why that 12:29:18
is in that equation. 12:29:18

Q. Is this supposed to be an effect of smoking apart 12:29:26
from its biological effect? Is that what we're 12:29:30
measuring here when you use the word demand? 12:29:32

A. That could be one possibility. 12:29:34

Q. What else could be in there? 12:29:34
A. I'm not sure right now. I actually find it 12:29:40
difficult to think about the equation without seeing 12:29:42
it in front of me, and I will review it at lunch. 12:29:46

Q. Well, I know you have big lunch plans, and I don't 12:29:50
-- we can come back to that tomorrow. I don't want 12:29:54
to get in the way of lunch. I'm serious, if you can 12:29:58
get it in mind at lunch, that's fine, otherwise we 12:30:02
can get it the next day. 12:30:04
Do you know how -- you probably don't 12:30:42
remember this, either -- how the pubsmk or privsmk 12:30:50
variable was figured, whether it's a 01 or some 12:30:52
other kind of construct? 12:30:54
A. See, I think it was possible for people to shift 12:31:06
their source of payer. And it may be that it's 12:31:12
reflecting the fraction of time that -- my 12:31:16
recollection is it may be reflecting the fraction of 12:31:20
time that they were receiving payment from one 12:31:22
source or another, but I don't remember exactly. 12:31:26
Q. Now, we're going to take the information from our 12:31:48
conditional probabilities and our expense regression 12:31:54
and we are then going to calculate the expected 12:32:06
expense in three different scenarios, nonsmoker 12:32:18
without disease, nonsmoker with disease, and smoker 12:32:22
with disease, correct? We need that for our final 12:32:28
equation? 12:32:30
A. We need three terms, expenditures for smokers with 12:32:34

disease, for nonsmokers with disease, and for 12:32:36
smokers with -- I'm sorry, for nonsmokers without 12:32:42

disease. 12:32:42
Q. And jumping ahead a moment, we're going to -- 12:33:10
everything we've done and talked about so far is 12:33:12
being done in NMES, just to get us grounded, 12:33:16
correct?
A. The estimation of these models is done with NMES 12:33:20
data, correct. 12:33:22
Q. Now, ultimately we're going to run people from BRFSS 12:33:34
through these models, is that a term that makes 12:33:40
sense to use it that way? We are going to apply 12:33:42
these models to BRFSS persons. How do you say it? 12:33:46
A. Behavioral Risk Factor Survey. 12:33:50
Q. Yeah, but am I saying that right, we are going to 12:33:54
apply these models to the BRFSS people, or we're 12:33:56
going to run the BRFSS people through these models? 12:33:58
A. I wouldn't say it that way. 12:34:00
Q. How would you say it? 12:34:02
A. I would say that we're going to apply these models 12:34:04
to the claims data, and that where information is 12:34:14
lacking on the claims data about certain variables 12:34:16
in this model, we will assume values for those 12:34:24
variables like we see for people in Minnesota, as 12:34:30
assessed in the Behavioral Risk Factor Survey. 12:34:32
For example, information on overweight or 12:34:32
seat belt use. 12:34:36

THE WITNESS: I don't mean to interrupt 12:34:38
the proceedings, I would very much appreciate a 12:34:40
chance to use the bathroom now. 12:34:56

THE VIDEOGRAPHER: Temporarily going off 12:35:00
the record. The time is 12:35 p.m. 12:35:04

(A lunch break was taken.) 12:35:06

THE VIDEOGRAPHER: We're back on the video 13:56:32
record. This is the third tape of the videotape 13:56:34
deposition of Scott Zeger. The time is now 1:56 13:56:38
p.m. 13:56:38

BY MR. SILFEN:

Q. Well, we're back from lunch. Hamlin and Zeger had a 13:56:42
big lunch out. I sat here and thought every 13:56:46
minute. 13:56:48

Here was my thought, in the refined 13:56:54
disease model for lung cancer/COPD, the only factors 13:57:02
you controlled for are age, gender, disease? 13:57:08
Anything else, that's it? 13:57:10

A. That's it. 13:57:10

Q. For CHD/stroke, you controlled for a number of other 13:57:18
factors? 13:57:20

A. Correct. 13:57:20

Q. And forgive me if I asked you this before, but I'm 13:57:26
not sure I did. Why that difference? Why you 13:57:30
controlled for a number of factors for CHD/stroke 13:57:34

but not for lung cancer? 13:57:34
A. This was recommended by Dr. Samet, and it was, as I 13:57:38
recall his rationale, it was because of the fact 13:57:46
that COPD and lung cancer have such a very high 13:57:50
attributable fraction, that is to say that smoking 13:57:54
is the very much dominant reason for the biological 13:57:58
process that leads to those diseases. 13:58:00

Whereas for the others, it wasn't to the 13:58:02
same extent. And that other factors, such as 13:58:04
socioeconomic status and the like might play a 13:58:08
relatively more substantial role than they do for 13:58:12
lung cancer in COPD. 13:58:14

Q. But what would have been the cost of using the same 13:58:22
methodology and controlling for other factors? I 13:58:28
guess why not do it, especially since so many 13:58:32
factors were dropped out if they didn't need to test 13:58:36
their significance, anyway? 13:58:38

A. It was a judgment call. Typically when there is a 13:58:44
strong biologic rationale for making a decision on a 13:58:48
model, I think that tends to be followed and there 13:58:52
is a cost of having to estimate more things, time, 13:58:58
you know, computer runs, variability added to the 13:59:02
estimate when variables, which aren't necessary, are 13:59:06
added to a model. So I think those were the 13:59:08
considerations that led us to this decision. 13:59:10

Q. Isn't it an unconservative way to do it? 13:59:12

A. I don't know if it's unconservative or 13:59:16

anticonservative or some other conservative. 13:59:20
Q. Here actually was my thought over lunch. Does that 13:59:28
rationale hold up when you have made another choice, 13:59:34
which is to include in the dollars the heart 13:59:40
disease, diminished health status, and every other 13:59:44
expense incurred by those people? 13:59:46
A. Yes. 13:59:48
Q. Why? Let's assume for the moment that it's not 13:59:52
trivial, that 30 or 40 percent of the dollars that 13:59:56
we are putting in the lung cancer pot are not lung 14:00:00
cancer dollars. Wouldn't we want to take other 14:00:06
factors into account? 14:00:06
A. Well, I've already answered that there's a couple 14:00:18
things. First is that the dollars which are not 14:00:20
lung cancer dollars might be as large as they are 14:00:24
because of the presence of lung cancer. 14:00:32
There is still the problem of having to do 14:00:36
the estimation when -- if you go ahead and try to 14:00:46
fit the model for the lung cancer/COPD as you have 14:00:48
for the others. 14:00:50
And there is -- I would grant the 14:00:54
possibility that if what we have in the lung 14:01:04
cancer/COPD is a mix of -- well, is an addition to 14:01:08

lung cancer/COPD some other costs, and if those 14:01:10
costs are strongly associated with these controlling 14:01:14
variables, then you might actually get some ability 14:01:16
to do some control from those additional costs. 14:01:18

That, I would agree with. 14:01:20
Although, I would add that the difference 14:01:24
between the core estimates, which controlled only 14:01:28
for those I think 16 strata, and the refined model 14:01:34
was not very large, the refined model having 14:01:38
controlled for all those other factors. 14:01:40
So I would speculate at this moment that 14:01:44
if we had done that, it would be unlikely to have 14:01:46
made much difference since doing all the control for 14:01:50
the other sorts of things we looked at didn't seem 14:01:52
to make very much difference. 14:01:54

Q. Of course, the principal point we're making here, 14:01:58
which is the lung cancer wasn't just lung cancer 14:02:00
dollars, applies to the core model, as well? 14:02:02

A. I've already addressed, though, that issue. 14:02:06

Q. Well, that was my lunchtime thought. What was your 14:02:10
lunchtime thought? 14:02:12

A. I thought a little bit -- well -- 14:02:14

Q. Couldn't trick you. 14:02:16

A. I thought about tuna fish. 14:02:18

Q. Okay. I tell you what, I looked up Dr. Miller's 14:02:34

answers to the questions we were talking about 14:02:34
here. Let me give you copies of his deposition, 14:02:44
which I know you've already seen. I lugged them all 14:02:48
the way from Washington. Page 209. Actually page 14:02:54
208, 209. It's the first page of the second day. 14:03:00

A. I don't have page 208 or 209. 14:03:04

Q. You probably do, actually. 14:03:06

A. Do I? 14:03:06

Q. Oh, it's one fat copy. 14:03:18

MR. HAMLIN: This is Volume I. This is Volume II. 14:03:28 14:03:28

BY MR. SILFEN:

Q. Bottom of 208, it's me saying, "In the refined disease model," do you have that? 14:03:34 14:03:36

A. Line 25, "In the refined disease model in the"? 14:03:38

Q. "Expense regression," and then I say, "are we in the same place?" That's a routine you're familiar with, "Mm-hmm (yes)." 14:03:42 14:03:44 14:03:44

A. Right.

Q. "I asked whether the disease variable was disease specific or was yes, no, any smoking-related disease, and the answer is? Any." 14:03:48 14:03:48 14:03:50

Do you agree with that? 14:03:54

A. Yes, I do. 14:03:54

Q. And also I had actually earlier asked him why it was 14:04:04

"any" rather than "disease specific," but his first instinct, like yours, was that it was "disease specific." And he answered that at page 137. 14:04:08 14:04:10 14:04:18

A. Okay. 14:04:28

Q. And you see me saying, "I may well have asked already, which is why am I seeing an any disease coefficient rather than a COPD/lung cancer coefficient and a CHD/stroke coefficient." 14:04:32 14:04:36 14:04:38 14:04:44

And if you'll read through there, you'll 14:04:46

see that Dr. Miller's recollection was that you 14:04:50
would try it and that it didn't make any 14:04:54
difference. 14:04:56

A. I don't have any other -- 14:04:56

Q. Explanation? 14:04:58

A. To add to that. I don't know. 14:05:04

Q. Well, let's take a look at his last line there. I 14:05:08
said to him -- 14:05:10

A. Give me the page again. 14:05:10

Q. 137. 14:05:12

A. Okay. 14:05:12

Q. Line 17. I said, "In other words, given that you 14:05:16
had a one, any one of the major smoking-related 14:05:22
diseases, there will be no difference in the 14:05:24
expected expenditure?" And his answer was, "Data 14:05:26
was unable to distinguish between one classification 14:05:28

and another." 14:05:28
Does that jog your recollection? 14:05:30

A. No, I have no other recollection about this. 14:05:32

Q. Okay. Let's assume Dr. Miller is correct. What is 14:05:40
the significance of your regression being unable to 14:05:44
distinguish between the expense of one 14:05:46
smoking-related disease and another? 14:05:48

A. I know of no special significance. 14:05:56

Q. Well, do you think that it's a fact that all 14:06:00
diseases have the same cause? 14:06:02

A. I would assume it's not a fact that all diseases 14:06:10
have the same cost, but I don't know the extent to 14:06:14

which they differ one from the other. 14:06:14
Q. Well, I assume with you that all diseases do not 14:06:18
have the same cost. And so I ask again: What does 14:06:22
it mean this regression can't tell the difference 14:06:26
between one disease and another? 14:06:28
A. I don't think it means anything. 14:06:28
Q. Okay. To your knowledge, does given that you have a 14:06:38
smoking-related disease or one of the major smoking 14:06:42
attributable diseases, does it matter to this report 14:06:46
whether you're a smoker or a nonsmoker? 14:06:48
A. When you say to this report, what do you mean? 14:06:54
Q. Yeah, I mean, anywhere in it. Well, we can stick to 14:06:58
this expense regression, if you want. But to the 14:07:00

expense regression or the average expense, do you 14:07:02
know if it matters whether you are a smoker or a 14:07:06
nonsmoker? 14:07:06
A. Well, the models for the major smoking-related 14:07:10
diseases estimate the expected expenditures for 14:07:16
smokers versus for nonsmokers or for -- and so in 14:07:20
that sense it does matter, yes. 14:07:22
Q. But the average expense, if you have a 14:07:24
smoking-related disease, does it -- 14:07:28
A. Or if you don't have a smoking-related disease. 14:07:28
Q. Does it differ between a smoker and a nonsmoker? 14:07:34
A. I'm sorry, ask that question again. 14:07:34
Q. Does the average expense, expected expense, given 14:07:38
that you have one of the smoking-related diseases, 14:07:40

differ between a smoker and a nonsmoker, do you 14:07:44
know?

A. Are you asking about the fit of the model to the 14:07:48
data, the coefficients in the model? 14:07:50

Q. I'm asking whether in this model, okay, a smoker 14:07:54
with a smoking-related disease has a different cost 14:07:58
than a nonsmoker with a smoking-related disease? 14:08:02

A. I believe that this model allows for that 14:08:04
possibility. 14:08:04

Q. Okay. All right. Well, then we'll find out where. 14:08:14
Now, one thing, when we broke we were 14:08:20

talking about how the BRFSS people are ultimately 14:08:24
going to be used when we get through this model. 14:08:26
Am I correct that first we estimate the 14:08:38
model wholly in NMES, correct? 14:08:46

A. Correct. 14:08:46

Q. And then we create a SAF. And the SAF is a product 14:08:52
of the parameters of the model estimated in NMES and 14:08:58
the characteristics of the BRFSS population? 14:09:06

A. Not entirely. 14:09:08

Q. What else is in there? 14:09:08

A. The characteristics of the group of people for whom 14:09:12
we have a pot of money. 14:09:16

Q. How do they affect the SAF? 14:09:22

A. Well, we know characteristics. We basically started 14:09:28
with the expenditure data for the State of Minnesota 14:09:28
and Blue Cross/Blue Shield; we have every 14:09:30
expenditure. 14:09:30

And we break it up into little smaller 14:09:32
pots for which we know the person's age, we know 14:09:36
their gender, we know their marital status, often we 14:09:42
know their education, we know things from the 14:09:44
billing record. 14:09:46

And so we have the whole pool of money for 14:09:48
people like that with those variables. Those are 14:09:52
variables that are in our model we fit from NMES. 14:09:56

Now, there are some additional variables, 14:09:58
as well, in the model from NMES, that we fit from 14:10:00
NMES, like seat belt use and like whether they're 14:10:02
overweight. 14:10:04

Q. But surely you are not telling me that the 14:10:06
information in the claims data affects the 14:10:14
calculation on percent? 14:10:16

A. I am telling you that. 14:10:18

Q. How is it used? The calculation of the SAF, not the 14:10:22
smoking attributable expense that you ultimately 14:10:26
calculate. 14:10:26

A. Yes, I heard what you said. 14:10:28

Q. How does it -- we looked together at the formula, at 14:10:32
the ultimate formula. 14:10:32

A. Yes. 14:10:32

Q. Can you show me in the formula where the information 14:10:38
from the claims data enters the formula? And I'm 14:10:44
not asking you where factors, demographics like 14:10:48
those in the claims data are used in the formula, 14:10:52

I'm asking you where the information in the claims 14:10:54
data is used in the formula? 14:10:56

A. I don't understand your question. 14:10:58

Q. Well, then show me where the information in the 14:11:00
claims data is used in the formula. 14:11:02

A. Well, the claims data includes information like the 14:11:04

person's gender or like their education I think in 14:11:10
some cases. So where it's used is in equation 3.1 14:11:16
from this document 2324. 14:11:24

It has six different quantities, which are 14:11:26
estimated from our bivariate probit model, and from 14:11:32
the expenditure, any expenditure and size of 14:11:36
expenditure. Those models are fit to the data. 14:11:38

Q. Right.

A. And the predictors for those quantities include 14:11:44
things like education, which is on the claims form. 14:11:48

Q. Right, but -- 14:11:50

A. And so we take that education value and we stick it 14:11:54
in the equation, which allows us to calculate these 14:11:56
six quantities with the quantities we need to make 14:12:02
the SAF calculation. 14:12:04

Q. So, for instance, are you telling me that if the 14:12:08
education -- do individuals from the claims data get 14:12:18
a SAF? 14:12:18

A. No, I did not say that. 14:12:20

Q. Do individuals from BRFSS get a SAF? 14:12:24

A. Well, individual records, perhaps -- yeah, 14:12:32
individual records from the Behavioral Risk Factor 14:12:36

Survey are matched to the category, the claims 14:12:42
category, matched by age, by gender, and so forth. 14:12:48
And the reason we do that is because we 14:12:48

then get some additional information about people 14:12:52
for whom -- people in that claims for whom we're 14:12:54
trying to calculate a SAF. 14:12:58
And that additional information is 14:12:58
information like whether they're overweight and so 14:13:02
forth. And what we do then is we calculate the SAF 14:13:06
using these six quantities which we've talked 14:13:10
about. 14:13:10

Q. Sure. 14:13:10

A. Using the demographics from the claim, because 14:13:14
that's the people whom we're going to apply the SAF, 14:13:16
and the additional information we were able to get 14:13:18
from BRFSS, which we otherwise wouldn't have because 14:13:22
it's not on the claims data.

Q. Is it your position that the proper pronunciation is 14:13:26
BARFUS (phonetic)? 14:13:28

A. I apologize, it's the Behavioral Risk Factor 14:13:28
Survey. I actually have difficulty with either
BRFSS or BARFUS (phonetic). 14:13:32

Q. We have to get the important stuff out. 14:13:38

A. I appreciate the subtle contribution you made by 14:13:42
using BRFSS rather than BARFUS (phonetic). 14:13:42

Q. You're telling me something I didn't know, and 14:13:46
that's why I'm giving you trouble on this. I 14:13:48

thought, my impression was that it was BRFSS 14:13:52

demographics that were used to weight the parameters 14:13:58
of the model. 14:14:00

You got a SAF, then the SAF was applied to 14:14:04
the claims pot, a pot that was made from the claims 14:14:08
data, but you're saying that's not correct? 14:14:12

A. I think we're actually saying the same thing, 14:14:16
because that demographic information is both on the 14:14:18
claims record and also on the Behavioral Risk Factor 14:14:24
Survey information. 14:14:24

Q. If what you're saying -- 14:14:26

A. They're matched that way. 14:14:26

Q. Okay. If what you're saying is that the claims data 14:14:30
is broken up into cells, I don't like your term 14:14:34
little pots, into cells -- 14:14:36

A. What do you think about big pots? 14:14:42

Q. -- into cells with particular sets of 14:14:46
characteristics, and then you take BRFSS people, or 14:14:50
BRFSS records, if you will, with similar 14:14:52
characteristics, and perhaps some additional 14:14:54
characteristics, and use them to weight the 14:15:00
parameters of model and then you get a SAF and then 14:15:06
you lay it back against that cell, then I understand 14:15:12
you. 14:15:14

Is that what you're saying? 14:15:14

A. Let me say it in my words to be explicit. 14:15:14

Q. Sure. 14:15:14

A. What we're doing is calculating a SAF to apply to 14:15:18
claims data, expenditure data. 14:15:22

Q. Right. 14:15:22

A. And we have done a regression analysis so that that 14:15:24
SAF can vary by factors, some of which are listed on 14:15:32
the claims data, and so we're going to use the SAF 14:15:36
that corresponds to those variables, demographic 14:15:38
variables, on the claims record, but some of which 14:15:40
are not. 14:15:40

And to get information about the ones that 14:15:42
are not, we take all the people in the Behavioral 14:15:46
Risk Factor Survey who match this cell, this cell of 14:15:50
claims with respect to the demographics, but who 14:15:54
also then have information about seat belt use and 14:15:56
overweight. 14:15:56

And we calculate the average SAF to apply 14:16:00
to this claims cell using the values -- we get the 14:16:06
average SAF by using the information about 14:16:08
overweight, seat belt use, and so forth, that's not 14:16:12
on the claim data, that is in the behavioral risk 14:16:14
factor data to apply to the cell that we're talking 14:16:18
about. 14:16:20

A. And -- 14:16:24

Q. Is every person in BRFSS given an opportunity to 14:16:32

have a SAF? 14:16:34

A. To the best of my knowledge, every person for whom 14:16:40
we have a -- every person for whom we have a record 14:16:44
on the Behavioral Risk Factor Survey who corresponds 14:16:46
to a claims cell gets to contribute their additional 14:16:54
variables so that the SAF reflects the average, the 14:16:58
SAF reflects the average of those additional 14:17:00
variables over the behavioral risk factor people in 14:17:04
that cell. 14:17:04

Q. What are the factors that you recall define the 14:17:22
cells? 14:17:22

A. Let's see, it's age, it's gender, it's obviously a 14:17:32
pair type. I believe not on all but on some of the 14:17:36
records it also includes marital status, educational 14:17:42
level. 14:17:42

Q. I think race, too. 14:17:44

A. I think race, as well. 14:17:46

Q. Dr. Miller said -- 14:17:46

A. I'm not sure of all of them. 14:17:48

Q. Dr. Miller described this in the context of the 14:17:50
diminished health status model, but I didn't -- so 14:18:02
when we say match up, do we mean that you're going 14:18:12
to take a BRFSS person and match him or her exactly 14:18:18
for age, gender, race, and the other factors you've 14:18:24
described with someone in a claims data cell? 14:18:32

A. Well, the cell is defined in terms of these 14:18:38
variables I've listed to the best of my 14:18:42
recollection. 14:18:42

Q. Right. 14:18:42

A. And we then find all of the people in the Behavioral Risk Factor Survey who would fall into that cell by virtue of their values for those variables. 14:18:48 14:18:52 14:18:56

Q. Does the cell have a tolerance band, or is it you are -- 14:19:02 14:19:02

A. We're not matching. 14:19:02

Q. -- you are a married person, you are Hispanic, you are -- well, is it age, one-year age specific? 14:19:08 14:19:16

A. I'm not sure. You'd have to refer to Dr. Miller, I think. I don't know exactly the width of the intervals. 14:19:20 14:19:24 14:19:24

Q. But whatever those factors are, you're going to have to be exactly on point. You're going to have to match them exactly? 14:19:26 14:19:30 14:19:30

A. However the claims cell is defined, it might be defined in terms of multiple years of age or in terms of multiple education categories. 14:19:34 14:19:36 14:19:40

But however it's defined, as I understand the application, the BRFSS people are chosen so that they match those defining variables. And then we calculate the SAF, the values of those variables 14:19:42 14:19:46 14:19:50 14:19:52

which define the cell, and the SAF is averaged over the other variables that we didn't have information on the claims record. 14:19:56 14:20:00 14:20:00

To be precise, when we say average, what we actually do, and the reason perhaps there's 14:20:02 14:20:06

some -- where there was a little bit of 14:20:06
miscommunication was we actually calculate the SAF 14:20:10
for all the people who fall into that cell and then 14:20:12
use the average of the SAFs rather than taking the 14:20:16
average of their behaviors, behavioral variables, 14:20:20
and then you plugging that in once to the SAF 14:20:22
calculation as a way to deal with some nonlinearity 14:20:26
that can arise if you do the latter. 14:20:28

Q. And you implied earlier, although I'm not sure you 14:20:32
said you knew, that there would be some BRFSS people 14:20:34
that would not be used because they didn't match 14:20:36
up? 14:20:36

A. I don't know that that would be the case. I didn't 14:20:40
do the matching myself, and I'm not sure. 14:20:42

Q. Peter asked whether two people could match with one 14:20:54
or was only one-to-one? 14:20:56

A. Match with one what? 14:20:58

Q. One person in a cell with another. 14:21:00

A. Well, I didn't -- the matching is not with 14:21:04
individuals in the cell, everybody in the cell. In 14:21:08

fact, what happens is we accumulate the dollars in 14:21:10
the cell. It really is a pot. We accumulate the 14:21:20
dollars, but the dollars have a label on them which 14:21:20
is these demographic characteristics. 14:21:22

Q. This may be a semantic point, but, nonetheless, 14:21:28
let's assume that everybody from BRFSS is used, 14:21:34
okay? In what sense does the Minnesota claims data 14:21:44
affect the SAF? 14:21:48

Haven't you done nothing more than create 14:21:54
compartments that your SAFs can be laid back up 14:21:58
against? 14:21:58

MR. HAMLIN: Objection to form. 14:22:02

THE WITNESS: Yeah, I don't understand the 14:22:04
question. 14:22:04

BY MR. SILFEN:

Q. Here's my point. Let's assume for the moment, which 14:22:06
I have a feeling is true, that every BRFSS person is 14:22:10
used, you're not just discarding some. Okay? 14:22:12

A. Okay. 14:22:14

Q. And I think I would have heard about it from 14:22:20
Dr. Miller or Dr. Wyant somewhere along the way if 14:22:24
BRFSS people were being thrown in, thrown out, so 14:22:26
assume every BRFSS person is being used. 14:22:30

In what sense is the SAF, the results or 14:22:36
the SAFs, the results, affected by the claims data? 14:22:40

A. Well, it's affected by the claims data in the sense 14:22:46
that the values put into the equations fit in this 14:22:50
model. 14:22:50

Q. But -- 14:22:52

A. The values for those demographic variables which 14:22:56
define the cells, those are the things that go in 14:22:58
here to calculate the SAF. 14:23:00

Q. I understand, Dr. Zeger, but suppose there were no 14:23:04
claims data, wouldn't you have exactly the same 14:23:06
SAF? 14:23:06

A. No. 14:23:06

Q. I mean, your NMES, your parameters of your NMES 14:23:14
model would be the same without the claims data, 14:23:18
right? 14:23:18

A. The parameters of the NMES model are fit from the 14:23:22
NMES data. 14:23:24

Q. They would be the same regardless of what was in the 14:23:26
claims data? 14:23:28

A. Correct. 14:23:28

Q. And the BRFSS people, assuming we used them all, 14:23:32
they don't change, either, based on what's in the 14:23:34
claims data, right? 14:23:36

A. The BRFSS, the behavioral risk factor data are the 14:23:42
behavioral risk factor data.

Q. And if we use them all, we are going to get a SAF 14:23:46

that is a product of two things, the demographics of 14:23:50
the BRFSS people, and the parameters from the NMES 14:23:52
model, right? 14:23:54

A. Wrong. 14:23:54

MR. SILFEN: I'm sorry. Do you see where 14:24:00
I'm wrong? I'm just trying to shortcut this if I 14:24:02
can. 14:24:02

BY MR. SILFEN:

Q. Peter is suggesting that maybe what I'm missing and 14:24:28
what you're saying is that the BRFSS data gets 14:24:30
somehow weighted by the creation of the cell, is 14:24:32
that it? 14:24:32

A. We calculate a SAF only for a cell. And the people 14:24:38

that we -- whose SAFs we average across are the 14:24:42
people who match that cell. 14:24:44

If there wasn't that cell, it wouldn't be 14:24:46
used. So how the people are grouped together 14:24:48
depends upon the claims data. 14:24:50

Q. Where I went wrong was I kept on thinking that each 14:24:52
BRFSS individual goes in. If that were true, then 14:24:56
what I was saying would be right. But if they go in 14:25:00
and they create only a cell SAF, I see the 14:25:02
difference. 14:25:04

So what's the -- I think I see that the 14:25:10
demographics of the claims data then do affect, do 14:25:14

play into the formula. 14:25:20

Okay. Well, that was very helpful. Most 14:25:26
of the time we hope we're finding out things we 14:25:28
know, but I didn't understand that. 14:25:30

Now, where we were when you guys took your 14:25:46
luxurious lunch was we were at the -- we were still 14:26:00
in the expense part of the model, and we had talked 14:26:02
about the estimation of the regression. 14:26:06

And now, as I understand it, the next 14:26:12
thing that happens is we have to take our -- some of 14:26:16
the information from the regression and some of our 14:26:20
demographics and some of our conditional 14:26:22
probabilities and we have to figure out an average, 14:26:26
an expected average expense, for three different 14:26:28
scenarios; is this correct? 14:26:32

A. You're referring to equation 3.1? 14:26:34

Q. Well, no, actually, I'm referring to components of 14:26:38
the equation. 14:26:40

A. The expenditure components of the equation 3.1? 14:26:44

Q. As I understand it, we have to calculate an average 14:26:46
expected expense for a nonsmoker without a current 14:26:48
treated disease and nonsmoker with a current 14:26:52
treatment and a smoker with current treatment? 14:26:54

A. I agree. 14:26:56

Q. And I actually think that the last -- this is the 14:27:02

last piece of this puzzle we're all happy to say. 14:27:06

I understand that there are -- the 14:27:10
calculation is different for each of those three 14:27:18
scenarios, but the components are these: First, the 14:27:24
coefficient for the various covariates we're going 14:27:34
to use, such as education. 14:27:36

And, second, for the smokers with and the 14:27:50
nonsmokers with, the coefficient for the disease 14:28:00
covariate in the regression model, you with me so 14:28:06
far? 14:28:06

A. I think so. 14:28:06

Q. And those are easy. And then we also have for the 14:28:26
nonsmoker without, we have the coefficient for 14:28:34
pubsmk if not significant multiplied times the 14:28:44
percentage of smokers in BRFSS. You're looking -- 14:28:50

A. You lost me at the end there. 14:28:52

Q. Okay. I have four pieces of the equation. The 14:28:58
first two were easy; the coefficient for the 14:29:02

demographics and the coefficient for the disease factor, okay. 14:29:06 14:29:08

There are two others. They are both the pubsmk or privsmk factor. For nonsmokers without, you get this term, coefficient for pubsmk if not significant and it is multiplied times the percentage of smokers from BRFSS. 14:29:14 14:29:26 14:29:40 14:29:42 14:29:44

And for smokers with -- the record should reflect the witness is already shaking his head, but why don't I finish and then you'll have it all and we'll talk about it. 14:29:52 14:29:54 14:29:56 14:29:58

For smokers with, we get also, we get the coefficient for pubsmk or privsmk if significant multiplied times the probability of smoke, given smoking-related diseases. 14:30:06 14:30:12 14:30:20 14:30:26

Just all blank here? 14:30:28

A. I can't tell you whether that's correct or not. 14:30:30

Q. And I take it you also don't have any sense what it means? 14:30:34 14:30:34

A. It was a long stream of words. It's hard for me to follow. 14:30:40 14:30:40

Q. I understand. Why don't I back up and make it simple. Let's assume that in the equation for the calculation of the average expense we have our demographics, our coefficient for demographics and our coefficient for disease, yes, no, and now we have a pubsmk or privsmk term? 14:30:42 14:30:46 14:30:48 14:30:52 14:30:56 14:31:00

A. Agreed. Let me see if I -- what I can agree. The 14:31:06
regression coefficient, the regression model, both 14:31:08
for the probability of any expenditure and for the 14:31:10
size of the expenditure, given any, that involves 14:31:14
the covariates, smoking, whether it be pub or 14:31:18

private, and disease, currently-treated disease. 14:31:24

Q. So we have those three parts. Dr. Miller also did 14:31:30
not at first asking could not explain this. And 14:31:34
then the next day he did explain it. And I'm not 14:31:38
trying to do a test here. It was extremely puzzling 14:31:42
to us. And I would be interested in your 14:31:44
explanation of this construction. 14:31:50

I should -- you might want to look at his 14:31:52
deposition because he actually on the second day 14:31:54
came in and said that I was, in fact, reading this 14:32:00
from the code correctly, but that it was a mistake, 14:32:04
that you guys had -- that it wasn't what you should 14:32:06
have done. 14:32:08

I'm not saying he was saying it was an 14:32:10
awful huge thing. 14:32:12

MR. SILFEN: Do you want me to show him? 14:32:14

MR. HAMLIN: Sure. Well, it's your 14:32:20
deposition, Tom. I thought you were going to refer 14:32:22
him to a page, that's all. 14:32:24

MR. SILFEN: Maybe I should do that. 14:32:26

BY MR. SILFEN:

Q. If you look at page 210, and I don't know whether 14:32:36
I'll have you read it today or just you can take a 14:32:40

look at it and we can orient on what it is. 14:32:54
If you look at 210 right in the middle 14:32:56

you'll see me saying, "Okay. I asked in particular 14:33:00
why it appeared that pubsmk appeared twice and was 14:33:04
specified so that once it was in if not significant 14:33:06
and another place it was in if significant. Did you 14:33:10
see that?" 14:33:12
He goes on and explains, and if you follow 14:33:24
that explanation down to the bottom of 212, 14:33:28
Dr. Miller ends up saying that he's discovered an 14:33:32
error in the code last night when he looked at this 14:33:36
matter. 14:33:38
I think it would not be productive to have 14:33:42
Dr. Zeger read this now. But since you have no 14:33:44
recollection of this matter, if you could just take 14:33:48
a look at that page or two tonight. 14:33:50

A. I don't want it to appear that I have no 14:33:52
recollection of the matter. I was -- it's hard to 14:33:56
read an equation and, you know, be able to comment 14:34:00
whether it's correct or not, so I'd be happy to read 14:34:02
this. 14:34:04

Q. I think the record will reflect that you affirmed 14:34:06
there would be three pieces to this, which would be 14:34:08
the demographic covariates and the disease and the 14:34:12
smoking covariate, and that it was the details of 14:34:16
the smoking covariate that we had a problem. 14:34:20
Now, having said that, having said that, I 14:34:24

do -- I should also say that neither you nor 14:34:28
Dr. Miller were able to explain at first look what 14:34:32
seems to me to be a fundamental aspect of the 14:34:34
equation. So I will be looking forward to your 14:34:38
explanation tomorrow. 14:34:38

A. Well, I'm here to answer your questions. My comment 14:34:42
to you was that I didn't understand the question. I 14:34:44
feel capable to answer questions about this model 14:34:46
and I would be happy to do so. 14:34:50

Q. Well, then let me ask you again. Why is it that the 14:34:52
pubsmk or privsmk appear twice in the equation for a 14:34:58
smoker with once if specified if not significant 14:35:08
multiplied times percentage of smokers from BRFSS 14:35:12
and once specified if significant multiplied by the 14:35:16
probability of smoke given disease? 14:35:16

A. Are you referring to the smoking variable in the two 14:35:24
different regressions, the any expenditure and in 14:35:28
the size of the expenditure when you say 2? 14:35:30

Q. Yeah, it would be probability -- 14:35:32

A. The reason that there are two coefficients is 14:35:40
because another way to model this would be to have 14:35:42
one model for the log expenditures, for example, and 14:35:44
not have a separate model for any versus the size. 14:35:48

But often in data like this where a large 14:35:50
fraction of people have no expenditure -- 14:35:52

Q. This isn't the regression. I don't mean to interrupt you. This is not the regression. This is the computation of the average expense. We've already discussed the regression and what was in it. 14:35:54 14:35:58 14:36:00 14:36:02 14:36:04

Now we're talking about the calculation, the computation -- Peter also corrected me, it's not in there twice, once for probability and once for level. In the probability of expense calculation, computation -- 14:36:06 14:36:14 14:36:16 14:36:20 14:36:22

A. Probability of expense? 14:36:22

Q. Yes. For a -- 14:36:24

A. Probability of any expense you're saying? 14:36:28

Q. Yes. For a smoker with COPD, the factors, the pieces of that equation are the coefficient for your covariates, the coefficient for disease, and the coefficient for pubsmk if not significant multiplied times the percentage of smokers from BRFSS and also the coefficient for pubsmk if significant multiplied times the probability of smoke given disease? 14:36:38 14:36:42 14:36:48 14:36:52 14:36:54 14:37:02 14:37:06

MR. HAMLIN: Objection to form. 14:37:08

BY MR. SILFEN:

Q. I thought you had said you didn't understand this and we were -- 14:37:14 14:37:14

A. I didn't understand the question is what I said. 14:37:16

Q. What is this? 14:37:18

A. I don't know. 14:37:18

Q. Well, I've referred you to the pages where Dr. Miller looked at it, and it's only a page or two, so why don't you take a look tonight. 14:37:24 14:37:28 14:37:30

The other thing you might look at is I asked you whether a smoker and a nonsmoker would have a different disease, a different expense, given that they had a smoking-related disease. And you said that it was allowed for in the computation. 14:37:54 14:37:56 14:38:00 14:38:04 14:38:06

Do you remember that exchange? 14:38:08

A. I'm sorry, repeat that. A smoker and a nonsmoker both of whom had the disease? 14:38:12 14:38:14

Q. Would have a different expense for that disease in this -- 14:38:18 14:38:18

A. You asked, and I said I thought that was allowed for? 14:38:20 14:38:20

Q. Yes. And what I would like you to do, if you're going to look at this, is tell me where that's allowed for. 14:38:24 14:38:26 14:38:26

A. Okay. 14:38:26

Q. Given that we haven't got it straight, it's silly to ask that next question. 14:38:30 14:38:32

Let me ask you another question along the same line. Is there any place in the model, the 14:38:56 14:39:00

refined disease model, where the prevalence of smoking in BRFSS makes a difference? 14:39:02 14:39:08

A. This is the refined -- 14:39:14

Q. Anywhere, yes, in the refined, as opposed to the 14:39:18

prevalence in NMES. 14:39:20
A. -- refined disease model. To the best of my 14:39:22
knowledge, no. 14:39:22
Q. Just to round out how the BRFSS claims data people 14:40:24
are used, as I understand it, each BRFSS person or 14:40:36
record matched, now I understand, with a claims data 14:40:48
record, goes into the conditional probability 14:40:52
regression and his or her characteristics go through 14:41:06
those conditional probabilities, and the result is 14:41:10
for that person a probability of disease and a 14:41:14
probability of smoke. 14:41:14
Is that correct? 14:41:16
A. Well, if we take the demographic variables that 14:41:22
define the claims cell, and then take a person from 14:41:26
the Behavioral Risk Factor Survey that falls into 14:41:34
that cell, then by taking their other variables, 14:41:36
like whether they're overweight and so forth, we can 14:41:38
calculate the components of the SAF, all the 14:41:42
components in equation 3.1. 14:41:44
Q. Right. I understand that, and that is helpful. But 14:41:46
I was trying to get those components piece by piece, 14:41:50

and it seems to me there are two main components. 14:41:54
One is the, as there always have been, the 14:41:58
conditional probabilities and then the expense 14:41:58
piece. And they go into the conditional probability 14:42:02
regression -- 14:42:02
A. Right. 14:42:04

Q. -- with their demographics. 14:42:06

And the result is that for that person you 14:42:10
then have all the conditional probabilities based on 14:42:14
their demographics? 14:42:16

A. Correct. 14:42:18

Q. And along the way, what you have calculated for them 14:42:20
is a probability of disease and a probability of 14:42:24
smoke jointly? 14:42:28

A. You can calculate for -- I mean, for the set of 14:42:34
demographics variables and the set of variables that 14:42:38
come from a BRFSS record, you can calculate the 14:42:42
joint probability distribution of being currently 14:42:44
treated and smoking and, therefore, can calculate 14:42:48
the conditional probabilities used in the equation 14:42:54
3.1. 14:42:56

Q. And in a sense, each person -- I'll say each person 14:42:58
and I will mean by that what I think you mean 14:43:02
without trying to restate it, which is a combination 14:43:04
of the claims data and the BRFSS information. 14:43:08

Okay? That won't work? 14:43:10

A. I think I made it clear what I meant, yeah, and I 14:43:14
think we're on the same wavelength. 14:43:16

Q. You don't -- 14:43:18

A. I think we are. 14:43:24

Q. Now, we also put each person or record into the 14:43:34
expense equations and we put them in once as a 14:43:46
never-smoker without, once as a never-smoker with, 14:43:52
and once as a smoker with; is that correct? 14:43:56

A. I believe so, yes. 14:43:58

Q. Now, that means that we are, in a sense, going to 14:44:32
get a SAF for persons who are in reality 14:44:36
nonsmokers? 14:44:38

A. Well, I've tried in my answers to the last couple of 14:44:46
questions to make the following point: We're not 14:44:48
interested in a SAF for a person from the Behavioral 14:44:54
Risk Factor Survey, no interest to us. 14:44:56
What we're interested in is a SAF for a 14:45:00
cell of expenditures that's defined in terms of 14:45:04
demographic parameters. 14:45:06
But in an effort to do the best possible 14:45:10
job controlling for other factors that were thought 14:45:14
by Dr. Samet to be of interest, which were available 14:45:18
in the Behavioral Risk Factor Survey, in order to do 14:45:22
the best possible job, we found people from the 14:45:24

Behavioral Risk Factor Survey who matched this 14:45:28
expenditure cell. 14:45:30
And we calculated one SAF for that cell by 14:45:34
getting the components of the SAF for all the people 14:45:38
who matched and getting an average SAF from them. 14:45:42
We're not interested in a SAF for this person from 14:45:44
BRFSS who smokes -- we don't care. 14:45:48
We are just trying to use the BRFSS data, 14:45:48
the Behavioral Risk Factor Survey data, to complete 14:45:50
the story with respect to variables like overweight 14:45:54
and seat belt use that we had available to us in 14:45:58

NMES and, therefore, put into our regressions for 14:46:02
the possibility that we could control from them.

Q. Nonetheless, the claims data doesn't have smoking 14:46:06
information, right? 14:46:06

A. Correct. 14:46:06

Q. And as you've said, you don't use the BRFSS smoking 14:46:18
information, either? 14:46:18

A. Correct, in this disease model. 14:46:22

Q. Right. And, therefore, even if we take the view 14:46:32
that we're getting a SAF for the cell, you have 14:46:38
people contributing to that SAF who were nonsmokers 14:46:42
treated as smokers in the model; isn't that right? 14:46:48

A. We have information from the Behavioral Risk Factor 14:47:00
Survey people who match based upon demographics our 14:47:04

cell, some of them may be smokers, some not, I don't 14:47:08
know. You know, I'm assuming that that's the case. 14:47:12
But it's the average SAF which we 14:47:16
calculate for all those people using this equation, 14:47:18
as I've described, which we then applied to the 14:47:22
cell. 14:47:22

Q. To calculate an average, you had to have as an 14:47:26
intermediate step a value for each person that 14:47:30
becomes part of the average, right? 14:47:32

A. Yes. 14:47:32

Q. And some of those people in the intermediate step 14:47:36
are persons who did not smoke who, nonetheless, have 14:47:40
a SAF? 14:47:40

A. I don't disagree with your description of the 14:47:44

calculation. What I'm trying to make clear is that 14:47:50
we're not in the business of calculating a smoking 14:47:54
attributable fraction for a person on the Behavioral 14:48:00
Risk Factor Survey. So when you say they have a 14:48:00
SAF, I don't know what to make with that. I'm 14:48:02
trying to disagree with that.

Q. I will stay with your agreement with the 14:48:04
calculation. 14:48:06

What is the interpretation of a SAF for a 14:48:10
person who never smoked, in your intermediate step? 14:48:14

MR. HAMLIN: Objection to form. 14:48:18

THE WITNESS: Are you asking me about the 14:48:22
calculation of a -- 14:48:24

BY MR. SILFEN:

Q. You've told me that in your intermediate step each 14:48:28
of the persons, including people who do not smoke, 14:48:32
would have a SAF. 14:48:34

MR. HAMLIN: Objection; mischaracterizes 14:48:38
his testimony. 14:48:38

THE WITNESS: I did not say that. My 14:48:38
answer was to the contrary. 14:48:40

BY MR. SILFEN:

Q. I thought you said you agreed with the calculation? 14:48:44

A. You misrepresented what I said. 14:48:46

Q. Oh, my heavens, I certainly didn't mean to. 14:48:50

A. Good.

Q. Now, let's replay this. I said that if you've 14:48:58

gotten an average, then as an intermediate step you 14:49:06
must have had a SAF for each of the persons. And I 14:49:12
thought you said, "I agree with that as a 14:49:14
calculation." Am I wrong? 14:49:16

A. What I said is that we are calculating a SAF for a 14:49:24
claims cell, that we'd calculate that SAF by taking 14:49:32
the characteristics of individual persons in the 14:49:36
Behavioral Risk Factor Survey. 14:49:38

Those characteristics, most of which match 14:49:40

the demographics that define the cell plus those 14:49:44
additional ones, and we calculate a SAF with those 14:49:48
values for those variables. 14:49:50

Okay, that SAF doesn't necessarily belong 14:49:54
to that person. It's a SAF that uses the couple 14:50:02
variables from that person. 14:50:02

And then we average those values over all 14:50:02
the people and to obtain a SAF which we then apply 14:50:04
to the cell. That's what I said. And if you want 14:50:06
to consider this -- well, that's what I said. 14:50:14

Q. For expense, to get an expense difference between a 14:51:24
smoker and a nonsmoker, you have to compare a smoker 14:51:28
and a nonsmoker, correct? 14:51:30

A. We calculate differences in expenditures for 14:51:36
groups. We take the -- 14:51:38

Q. All right, a group of smokers and a group of 14:51:40
nonsmokers. You have to have a group of nonsmokers 14:51:44
to compare to a group of smokers to get a 14:51:46
difference, right? 14:51:48

A. Correct. 14:51:48

Q. In the expense model, the expense part of the refined disease model, what group of smokers do you compare to nonsmokers? 14:51:52 14:51:54 14:51:56

A. The data from NMES, using the data from NMES. 14:52:00

Q. In the expense model? 14:52:04

A. Yes. 14:52:04

Q. The data from NMES? 14:52:10

A. Yes. 14:52:12

Q. What group of smokers and nonsmokers from BRFSS do you use to weight the expense calculation in the model or from the claims data? 14:52:34 14:52:42 14:52:46

A. I'm sorry, could you read that back? I didn't quite get it. 14:52:52 14:52:54

(The requested portion read back.) 14:53:06

THE WITNESS: That's not the use we make of the behavior risk factor data in the disease model. 14:53:14 14:53:20 14:53:22

BY MR. SILFEN:

Q. Now, the BRFSS people had their own smoking information, why didn't you use that? 14:53:36 14:53:40

A. The information that was necessary to calculate a SAF using equation 3.1 in the document you gave me requires us to know something about the joint distribution of currently treated disease, smoking, and expenditures. 14:53:56 14:54:02 14:54:04 14:54:10 14:54:10

We were able to estimate that joint 14:54:12

distribution in NMES. We could not estimate that 14:54:16
joint distribution in the Behavioral Risk Factor 14:54:18
Survey.
Q. That assumes that you're going to use exactly the 14:54:20

model you already specified. My question doesn't 14:54:24
assume that. 14:54:24
My question is: Why did you specify a 14:54:26
model that required you to make up smoking data for 14:54:30
the BRFSS people? 14:54:30
A. We did not make up smoking data for the BRFSS, for 14:54:36
the Behavioral Risk Factor Survey data. We used the 14:54:40
model that was the best statistical practice in 14:54:44
order to estimate the quantity we wanted to know, 14:54:48
the smoking attributable expenditures. 14:54:52
And in order to do that, you needed to 14:54:54
know something about the joint distribution of 14:54:56
treatment, smoking and expenditure, which was 14:54:58
available in NMES but not in the behavioral risk 14:55:02
factor. 14:55:02
Q. Tell me again why this is the best statistical 14:55:04
model, this one you have here? 14:55:06
A. Because it allows us to calculate the smoking 14:55:08
attributable fraction, which is what we need to 14:55:10
apply to the expenditures in order to obtain the 14:55:14
smoking attributable expenditures. 14:55:16
Q. We had a discussion this morning which we, I think, 14:55:18
established NMES has the expenditure data you need 14:55:22
and the covariates that you need. 14:55:26

And I'm not sure you were able to explain 14:55:28

to me then why you couldn't just estimate a model 14:55:32
which had expenditure as its outcome and smoking and 14:55:38
other covariates as the predictors, you wouldn't 14:55:40
have had to make up this data. Why not? 14:55:40

A. The answer I gave you this morning was that I 14:55:48
couldn't respond to your hypothetical model that you 14:55:54
were sort of describing with lack of sufficient 14:56:02
detail for me to on the spot give a critique of it. 14:56:04

I tried to answer your questions about 14:56:06
what we did and why our model enables us to address 14:56:12
the question we were asked to address. I did not 14:56:14
say that -- I did not comment upon any failing of 14:56:20
the model that we used or any other sort of 14:56:22
suggestion. 14:56:24

Q. I didn't suggest that you did. On the contrary 14:56:26
you -- I asked you a question why you didn't do it a 14:56:30
different way, and you answered in effect because we 14:56:32
did it this way and it's the best way. I think 14:56:36
that's how we got back where we were. 14:56:38

MR. HAMLIN: Wait, object on the grounds 14:56:42
that mischaracterizes Dr. Zeger's testimony. 14:56:44

MR. SILFEN: We'll read it. Perhaps it 14:56:48
did. And if it did, I apologize. 14:56:50

BY MR. SILFEN:

Q. Let's just confirm your understanding of how 14:56:54

covariates were used or not used in the model. 14:56:58

What was the basis on which your group of covariates was originally chosen? Why did you use -- why did you have that base group? 14:57:08 14:57:12 14:57:14

A. These were the covariates which Dr. Samet indicated, of the ones that were available to us from NMES, would be most important to attempt to control for. 14:57:18 14:57:20 14:57:26

Q. And you had no input on that at all? 14:57:32

A. It was a discussion. He was the medical expert. We tried to derive the choice of variables based upon medical understanding to the extent possible. 14:57:36 14:57:38 14:57:40

Q. Is there any distinction between the variables that Dr. Samet chose and ones that Leonard Miller had been using for five years? 14:57:46 14:57:52 14:57:52

A. Is there a distinction? 14:57:56

Q. Do you know? 14:57:56

A. I don't know. 14:57:56

Q. Now, the variables were kept or not kept depending on some tests of significance; is that correct? 14:58:12 14:58:16

A. My understanding is that variables were retained in the model in groups. 14:58:22 14:58:28

If one of a member of the group had a P value that was less than approximately .15, something like that I think is what Dr. Miller has indicated to me. 14:58:30 14:58:34 14:58:36 14:58:36

Q. Okay. So you -- the way you're describing that 14:58:40
sounds to me like you were not a part of the 14:58:42
decision-making on this issue? 14:58:42

A. I did not make the decision to include or drop out 14:58:48
variables. 14:58:48

Q. What is your thinking about the practice of keeping 14:58:56
variables that are not significant based on some 14:59:04
kind of a nonstandard test, if that's a way to say 14:59:12
it? Is it right, wrong, doesn't matter? 14:59:14

A. Can you be more specific? 14:59:16

Q. Well, the variables were not kept because they met 14:59:20
the standard 95 percent significance, they were kept 14:59:24
on some other basis which you say is roughly .15. 14:59:30
Why? What's the reasoning for that? 14:59:32

A. It's actually not the standard that you would retain 14:59:36
variables in a regression model if their P value was 14:59:40
less than .05, that would not be a typically 14:59:44
standard practice. 14:59:44

Typically a looser criterion is used to 14:59:48
admit variables to a regression model. Although, 14:59:52
there's no hard and fast rule about this. Some 14:59:56
people prefer to be a bit more liberal in it, and 14:59:58
others a bit more conservative in terms of admitting 15:00:02
variables. 15:00:04

Q. Some people would keep them all? 15:00:06

A. Some might keep it all. There's a trade-off of 15:00:08
precision and an error in the estimation, and there 15:00:12

isn't a hard fast rule one way or the other. 15:00:16

Q. Would it also be appropriate to make the decision 15:00:42
whether to keep a variable or not based on whether 15:00:46
it has an outcome on the -- has an effect on the 15:00:50
outcome on the SAF? 15:00:58

A. I'm not sure I understand the question. 15:01:02

Q. Well, it would be possible, for instance, to -- for 15:01:10
an insignificant variable to have an effect on the 15:01:12
outcome that you're predicting, right? 15:01:16

A. On the total smoking attributable expenditures, for 15:01:20
example?

Q. Right. So another way to decide whether to keep or 15:01:24
not keep a variable would be based on whether it 15:01:26
affects the outcome, for instance, the smoking 15:01:28
attributable expenditure, or, for instance, the 15:01:30
smoking coefficient? 15:01:32

A. That might be another way. 15:01:36

Q. And that was not done? 15:01:38

A. No. What was done is what I -- I mean, I'm only 15:01:42
reporting what Dr. Miller has told me. I think his 15:01:46
testimony would be most appropriate for that. 15:01:48

Q. That's fair. Would you agree that the most 15:02:16
important results in the expenditure models are the 15:02:20

estimates of the coefficients multiplying the 15:02:22
categorical measures of smoking history? 15:02:24

A. I'm sorry, say that again. 15:02:30

Q. I was just reading this sentence, footnote 30. 15:02:34

A. Certainly the coefficients for smoking in the 15:03:20

expenditure models are important variables. 15:03:22

Q. Well, this footnote says the most important 15:03:30
results. Do you agree with that or don't you? 15:03:32

A. I'm not exactly sure that's what's being contrasted 15:03:38
with, I mean, results sort of includes a lot of 15:03:40
things, and I don't know exactly how to. 15:03:42

I, unfortunately, didn't write that one 15:03:46
sentence so -- I would certainly agree that the 15:03:48
smoking coefficients are very important. And 15:03:52
whether it's the most along what transit of scale 15:03:56
I'm not sure, but it's very important, yes. 15:03:58

Q. Okay. Well, that leads back to the question we were 15:04:02
on. Why wouldn't the better test for keeping or not 15:04:08
keeping a variable being whether it affected the 15:04:10
smoking coefficient? 15:04:12

A. Well, that might be one criterion to consider. But 15:04:30
another criterion is the strength of evidence that 15:04:34
that variable is an important predictor, which is 15:04:38
the one that was used here. 15:04:40

Q. But as far as you know, no test was done to 15:04:52

determine whether omitted variables would or would 15:04:54
not affect the smoking coefficient? 15:04:56

A. So far as I know. Although -- but I didn't do 15:05:02
that. 15:05:02

Q. That's fine. 15:05:04

A. And when you say -- let me just be specific, we're 15:05:10
talking about variables that we had the data 15:05:12

available that were in the model, and we're talking 15:05:16
about whether to drop certain variables that didn't 15:05:18
achieve a certain level of predictive ability. 15:05:22
In that regard, I don't know of anything 15:05:24
that was done looking at their relationship to the 15:05:26
smoking variable. 15:05:28

Q. Did you ever examine the other variables that were 15:05:38
available in NMES that might have been used? 15:05:38

A. I've looked at the NMES forms. It's been a while, 15:05:46
but at some point I did look. 15:05:48

Q. Well, one that's in there is exercise. Why did you 15:05:54
choose not to put exercise in, especially since you 15:05:58
seem to view it as important to do a little test of 15:06:06
the significance of exercise in a footnote. In 15:06:10
fact, I think the footnote we just looked at. 15:06:12

A. Yeah, I don't recall the specific question about 15:06:18
exercise. I recall there was a single exercise that 15:06:22
I believe asked somebody to choose between one of 15:06:24

two alternatives that were a description of 15:06:26
themselves, something about exercising three times a 15:06:30
week for more than 20 minutes or something versus -- 15:06:34
it was another descriptor, I don't remember exactly 15:06:38
what it was. 15:06:38
But I remember it not being a very -- 15:06:44
well, there being some reservations about the 15:06:48
question is the first point. 15:06:48
The second point about the exercise 15:06:50
variable was there was some concern that it would be 15:06:56

what some call an endogenous variable, that someone 15:07:00
would certainly answer that question that they don't 15:07:04
exercise for 30 minutes every three times a week if 15:07:06
they were, they had lung cancer, for example, or if 15:07:08
they were otherwise sick with a smoking attributable 15:07:14
disease. 15:07:14

So I think those were the two major 15:07:16
concerns about that variable. 15:07:16

Q. Well, my question really goes more along the lines 15:07:20
of some of the other questions I've asked. You have 15:07:24
exercise there in your NMES data, why wouldn't you 15:07:28
use it? You would have to agree it is a 15:07:32
well-recognized health factor, true? 15:07:34

MR. HAMLIN: Objection; asked and 15:07:36
answered. 15:07:38

THE WITNESS: Yeah, I believe I've -- 15:07:40

BY MR. SILFEN:

Q. I'm sorry if you already answered that. But why 15:07:44
wouldn't you just put the exercise -- 15:07:48

MR. SILFEN: I see why you're objecting, 15:07:50
Tom. I'm meaning to put another slant on it. 15:07:52

BY MR. SILFEN:

Q. I understand you're saying there was a concern it 15:07:54
might be endogenous, but I'm not sure why you 15:08:02
wouldn't put the variable in to see if it would 15:08:06
significantly affect the smoking coefficient? 15:08:08

MR. HAMLIN: Objection; asked and 15:08:12

answered. 15:08:14
BY MR. SILFEN:
Q. It hasn't been asked and answered. I'm asking, notwithstanding the objections, wouldn't it have been the conservative thing to do to put the exercise variable in and see what happens? 15:08:16 15:08:18 15:08:22 15:08:24
MR. HAMLIN: Same objection. Objection to form, as well. 15:08:26 15:08:28
THE WITNESS: There are a whole set of things that one might do, and we are doing many of them. This is not one that we did for the two reasons I explained. 15:08:32 15:08:36 15:08:40 15:08:42
BY MR. SILFEN:

Q. Diminished health status, are you considering dropping this part of the model? 15:09:10 15:09:12
A. Am I? 15:09:14
Q. Yeah. 15:09:14
A. No. 15:09:16
Q. The diminished health status model attributes $223 million to smoking. And I believe that one of our experts reported that $131 million of that was attributable to the pubsmk coefficient for males 19 to 34. 15:09:42 15:09:54 15:10:00 15:10:06 15:10:06
Were you aware of that? 15:10:08
A. No. 15:10:08
Q. Well, Dr. Miller was aware of that and said that was one of the things that your group was looking into, but you have no knowledge of that? 15:10:18 15:10:20 15:10:22

A. Well, we were looking into the -- I think a number 15:10:26
of things about the -- I'm trying to remember our 15:10:30
discussion. 15:10:30

It gets a little confused between the 15:10:38
different models, so I'm trying to think about what 15:10:40
we discussed about this one. 15:10:42

Q. Think how I feel? 15:10:44

A. Actually, I do recall a discussion about the point 15:10:52
that a large fraction, I don't remember the numbers, 15:10:54
but a large fraction of the expenditures came from 15:10:56

one of the groups. 15:10:58

And that was going to be something we were 15:11:00
going to follow-up on, and I'm not sure what the 15:11:02
status of that is right now. 15:11:04

Q. Okay. Let me briefly run through with you my 15:11:20
understanding of the diminished health status model 15:11:24
and see if I have it right. 15:11:26

If we put aside for a moment the sample 15:11:34
selection issue, the Mill's inverse ratio. 15:11:42

A. Okay.

Q. I see three basic regressions here. The first is in 15:11:54
the first the outcome is did a doctor ever tell you 15:12:00
that you had one of four smoking-related diseases; 15:12:06
is that correct? 15:12:06

A. I believe that's correct for the persons 35 and 15:12:14
older. 15:12:14

Q. And in that regression, there are a varied group of 15:12:28

covariates and smoke as a main effect. Are you 15:12:38
aware -- is that true? 15:12:38

A. I believe so. I don't have the equation in front of 15:12:42
me. My recollection is certainly smoking is in that 15:12:48
regression. 15:12:52

Q. And there is then a second regression in which the 15:12:54
outcome is self-reported, poor health status? 15:13:00

A. Correct. 15:13:00

Q. And in that regression we use a fitted value from 15:13:06
the first regression plus smoke and other 15:13:10
covariates. 15:13:12

A. For the 19 to 34-year-olds we used the reported 15:13:22
value of previous disease. For the people 35 and 15:13:24
older, we used the Emperical Bayes estimate of their 15:13:28
underlying variable, given what they told us on the 15:13:34
questionnaire. 15:13:34

Q. And then in a third equation, self-reported poor 15:13:46
health status becomes a variable predicting 15:13:50
probability and level of medical expenditure, along 15:13:56
with pubsmk or privsmk, correct? 15:14:04

A. That's my understanding. It's, again, the same sort 15:14:08
of -- it's the predicted value based upon both the 15:14:12
regression from the previous step, but also on their 15:14:14
reported value. 15:14:16

Q. For the self-reported poor health status? 15:14:18

A. Yes. 15:14:20

Q. What is the interpretation of the pubsmk coefficient 15:14:24
in that final regression? 15:14:28

A. It's the -- let's see, that final regression. 15:14:36
Q. The outcome would be medical expense? 15:14:40
A. Expenditure, yeah. 15:14:42
Q. Self-reported poor health status would be one 15:14:46
variable and pubsmk another? 15:14:48

A. Yeah, so pubsmk would be the effect of smoking on 15:14:56
public or for expenditures from the public domain, 15:15:00
having controlled for the person's, our best 15:15:04
estimate, of their health status, poor health 15:15:10
status, as well as the other factors. 15:15:12
Q. Why is the variable here pubsmk, but in the earlier 15:15:24
regressions it's smoke? 15:15:26
A. I think the decision was that because, again, of 15:15:32
issues related to demand and access to allow for the 15:15:38
possibility of a differential effect of persons 15:15:42
receiving public versus private insurance. 15:15:44
Whereas, the earlier equations are, again, 15:15:48
more directed at the biologic process for which that 15:15:52
was not considered as important, given the other 15:15:56
factors in the model. 15:15:58
Q. Why this structure? Why -- in this structure we 15:16:32
don't even have a direct input of smoking to medical 15:16:34
expenditure, why? 15:16:36
MR. HAMLIN: Objection; form. 15:16:40
THE WITNESS: I don't understand your 15:16:40
question. 15:16:40
BY MR. SILFEN:

Q. We specified a model here where we don't even have a 15:16:44
direct smoking factor as a predictor of medical 15:16:48
expense, it goes through self-reported health 15:16:50

status. Why? 15:16:52

A. I don't agree. 15:16:54

Q. You don't agree? 15:16:56

A. No. 15:16:56

Q. I take it you don't agree because of the pubsmk 15:17:06
variable? 15:17:06

A. Correct. 15:17:06

Q. And the pubsmk variable you've explained as having 15:17:18
an interpretation of a demand function above your 15:17:22
perceived health status, correct? 15:17:26

A. This model allows for smoking to work in a couple of 15:17:32
ways, either through self-reported poor health or 15:17:36
directly to expenditures. 15:17:42

Q. Yeah, but, once again, Doctor, the outcome you're 15:17:50
predicting is medical expenditures, correct? 15:17:54

A. Correct. 15:17:56

Q. Why wouldn't you just model directly from smoking 15:18:00
and the other covariates, including insurance, to 15:18:04
medical expenditure? 15:18:06

A. That would be another possibility. 15:18:10

Q. But why did you do it this way? 15:18:12

A. Because this model attempts to describe the process 15:18:18
by which expenditures might be elevated among 15:18:22
smokers. That process could reflect two pathways; 15:18:28
one which is more biologic in nature, and the other 15:18:32

is having more to do with their use of medical 15:18:34
services. 15:18:36

So this model is, as all good models, 15:18:42
mimicking the potential pathways, not simply being 15:18:46
descriptive. 15:18:48

Q. I take it, then, that if we modeled directly from 15:18:50
smoking and the other covariates to medical expense 15:18:52
we'd get the same result? 15:18:54

A. I don't know. I've not done that, so I can't say 15:18:56
you'd get the same result if I haven't done it. 15:19:00

Q. Why wouldn't you? And what would it mean if you 15:19:02
didn't get that result? 15:19:02

A. I don't know. I don't know. It's not something 15:19:04
I've done. I can't speculate about what would have 15:19:06
happened, and then can't tell you why it would have 15:19:10
happened if I don't even know what would have 15:19:12
happened. 15:19:12

Q. Did you ever sit down with someone saying why are we 15:19:20
doing three regressions here? Why do we do previous 15:19:28
disease and self-reported health status? Why don't 15:19:32
we just find out if smoking has a direct effect on 15:19:36
medical expenditures? 15:19:36

A. I did not ask that question, no. Let me also add 15:19:40
that you can find out what the effect of smoking is 15:19:48
on medical expenditures this way, which is why we 15:19:50

did it this way. 15:19:52

Q. Why wouldn't you have a smoke variable in the final 15:20:28
equation and a pubsmk variable, as well? 15:20:32

A. If I understand the question, you're asking why 15:20:44
wouldn't we have smoke plus pubsmk and private 15:20:50
smoke, is that the question? 15:20:52

Q. Yes. 15:20:52

A. It's not necessary. If you have -- if you define a 15:21:00
variable, which indicates public smoke -- and, 15:21:04
again, I think there's the fraction of time issue 15:21:06
that people were switching in and out of different 15:21:10
insurance status. 15:21:14
But let's just set that aside for the 15:21:16
moment. If you had one variable for private smoke 15:21:22
and another for public smoke, what you're doing is 15:21:24
allowing the effect of smoking to be different for 15:21:32
those two categories of insurance. 15:21:32
It's not necessary to have a smoking 15:21:36
variable -- well, in fact, it wouldn't be possible 15:21:36
if the smoking variable was the sum of the other 15:21:38
two, it would be colinear. 15:21:40

Q. Let me go back to the place we started. Let's 15:21:42
suppose that it's correct that in this model the 15:21:48
pubsmk coefficient for 19 to 34-year-olds is 15:21:54
responsible for more than half of the dollars. 15:21:58

What would that tell you? 15:21:58

A. It doesn't tell me anything in particular. 15:22:02

Q. Well, come on, I mean, what is the interpretation? 15:22:12
You have $231 million of smoking attributable 15:22:20
expense, and let's assume that it is correct that 15:22:24
the 19 to 34-year-old pubsmk coefficient is 15:22:28
responsible for 131 million of it. 15:22:32

What is that 131 million? How would you 15:22:34
interpret it? 15:22:34

MR. HAMLIN: Objection; asked and 15:22:36
answered, also on basis of form. 15:22:40

BY MR. SILFEN:

Q. $131 million of what? 15:22:42

A. U.S. currency, I assume. 15:22:44

Q. You know, you're the chief of biostatistics, and 15:22:50
this is not a silly question. I don't know how to 15:22:52
interpret that. I'm told that it's true. $131 15:22:58
million is for 19 to 34-year-old pubsmk 15:23:02
coefficient. This is your model. 15:23:04

I don't think it's out of line for me to 15:23:06
ask how he interprets that. And I didn't mean to be 15:23:08
facetious. 15:23:10

How do you interpret that? What does it 15:23:14
mean? 15:23:14

MR. HAMLIN: Objection; asked and 15:23:16

answered. 15:23:16

MR. SILFEN: I don't believe it was.

THE WITNESS: It doesn't mean anything. 15:23:18

It means that there's an estimate of $131 million 15:23:20
for the category that you named. That's the 15:23:22
estimate of the smoking attributable expenditures 15:23:26
for that category. 15:23:26

BY MR. SILFEN:

Q. Well, but I believe you said something to the effect 15:23:28
that this was the nonbiological effect of smoking. 15:23:32

Does that mean we have $131 million of 15:23:34
nonbiological smoking attributable effect? 15:23:36

A. It means for that category, the estimate of the 15:23:40
smoking attributable expenditures was $131 million, 15:23:44
just the way for some other category it was whatever 15:23:48
it was the other category. 15:23:48

Q. Yes, but this is a very specific category, this is 15:23:50
19 to 34-year-old, people for whom in previous 15:23:54
models you have ruled that they can't have a major 15:23:56
smoking-related disease, right? 15:23:58

A. They were -- they were not in either the lung 15:24:02
cancer/COPD or the CHD/stroke categories. 15:24:06

Q. So it's not a major smoking-related disease. You 15:24:08
said it's some kind of demand effect? 15:24:10

A. I named that as one of the things that it might be, 15:24:14

yes. 15:24:14

Q. How can you account for the fact that you'd have 15:24:20
$131 million of smoking attributable costs for 19 to 15:24:26
34-year-olds in this category? 15:24:30

MR. HAMLIN: Objection; asked and 15:24:32
answered. 15:24:34

MR. SILFEN: Tom, actually you may be 15:24:36
right. It is true I've asked it, but it seems to me 15:24:40
I'm looking for some -- I'm looking for one of the 15:24:44
creators of the model to give me some reason why 15:24:46
that's a logical result. And if I've had his 15:24:48
answer, then that's fine. 15:24:50

MR. HAMLIN: I mean, if you want me to get 15:24:54
into a speaking objection, I will. I think he has 15:24:56
answered the question. 15:24:58

MR. SILFEN: Okay. If that's the answer, 15:25:04
that's it. 15:25:04

BY MR. SILFEN:

Q. Do you know why there's two Mill's inverse ratios in 15:25:32
the diminished health status model? 15:25:34

A. In which equation? 15:25:40

Q. I don't know. All the equations, isn't it? Well, 15:25:48
I'm not sure. Are there two in some of the 15:25:50
equations? 15:25:52

A. Yes. 15:25:52

Q. In which equations are there two? 15:25:54

A. In the expenditure equation there's two. 15:25:58

Q. That's the one I'm referring to you as the third 15:26:00
where you're predicting medical expenditure? 15:26:02

A. To the best of my recollection, there are two. 15:26:04

Q. I try not to think about the Mill's inverse ratio 15:26:08
very much, but why don't you tell us quickly why 15:26:10
there are two, if you know? 15:26:12

A. You have to think of the hypothesized underlying 15:26:22 variables which describe your tendency to have poor 15:26:28 health, et cetera. 15:26:30

In this particular problem, we imagine 15:26:32 that there's such a variable for disease, as well as 15:26:38 for the sample selection process. 15:26:40

And so you want to calculate things like 15:26:44 the -- in these models like the expectation of what 15:26:50 you report for your health status, for example, 15:26:52 given that we know you're not currently treated and 15:26:58 given that we know that you were in the sample. 15:27:02

And so this model actually posits a 15:27:06 statistical amount for the joint distribution of the 15:27:10 three underlying variables, the health status, 15:27:14 tendency to be in the sample, and tendency to be 15:27:18 currently treated. 15:27:18

So the equation that's specified for, for 15:27:22

example, the poor health model is given that you are 15:27:26 not currently treated and given that you are in the 15:27:28 sample. 15:27:28

And so that conditioning -- actually, 15:27:34 these Mill's ratios are obtained by actually writing 15:27:36 down a model for the underlying variables and then 15:27:46 sort of integrating out some stuff, and they pop 15:27:48 into the equation because of formulas having to do 15:27:48 with those integrals. 15:27:48

So it's basically a selection process 15:27:54 that's gone on both in terms of being in the sample 15:27:56

and in terms of not being in the currently treated 15:27:58
group. 15:27:58

Q. I understand the -- 15:28:04

A. He'll ask tomorrow. 15:28:06

Q. I understand the sample bias issue because I had to 15:28:12
deal with it for a couple years. 15:28:16

The suggestion here is that there could be 15:28:26
a bias in the sample because you are a person not 15:28:32
currently treated, is that the fundamental notion? 15:28:34

A. I wouldn't put it that way, no. I know in the one 15:28:38
case we have a not being in the sample and now not 15:28:40
being currently treated. 15:28:42

I think an easier way to think about it is 15:28:46
that there are three underlying variables, your 15:28:48

tendency to be currently treated, your tendency to 15:28:52
be in the sample, and your tendency to have poor 15:28:54
health. 15:28:56

And the model that has been written down 15:28:56
is for the joint distribution of those three 15:29:00
underlying variables. 15:29:02

Now when we go to estimate the parameters 15:29:04
in one of those models, the poor health model, we 15:29:08
know something about the other two. 15:29:10

So the parameters which have been 15:29:12
formulated in terms of this underlying three 15:29:16
variables, we can't ignore the information we know. 15:29:20

The information we know is that you are in 15:29:20

the sample and you don't have a currently treated 15:29:26
disease. So when you say, okay, now I want to model 15:29:30
the expected poor health, given that I know those 15:29:34
two pieces of information, just write down the 15:29:36
equations and it turns out that you end up with the 15:29:40
Mill's ratio one for each of them. 15:29:42

Q. It's not your fault. Tell you what, let's take a 15:29:48
few minutes and then we'll go into a final stretch. 15:29:52

THE VIDEOGRAPHER: Temporarily going off 15:30:04
the video record. The time is now 3:30 p.m. 15:30:10

(A break was taken.) 15:30:10

THE VIDEOGRAPHER: Back on the record. 15:42:24

The time is now 3:42 p.m. 15:42:26

BY MR. SILFEN:

Q. Tell you what, we'll try a different procedure 15:42:28
here. You just tell me how you did the nursing home 15:42:32
model. 15:42:32

A. Let's see, the nursing home model actually uses the 15:42:42
same approach as everything else we've done, which 15:42:46
is to say we take total expenditures each year, 15:42:56
which in this case were Medicaid expenditures, and 15:42:58
we calculate a smoking attributable fraction. 15:43:02

And we use not the NMES but the NHANES, 15:43:10
the follow-up study of the NHANES study, to get -- 15:43:18
for the purpose of estimating the smoking 15:43:20
attributable fraction. 15:43:22

And we do that by taking all of the people 15:43:24
-- this period of follow-up we use is 1982 to 1992, 15:43:30

an 11-year period, and we take everybody in NHANES 15:43:34
who during that period turned 55 years old. 15:43:44

I think there are approximately 1,000 such 15:43:46
people, maybe nine hundred some people, about 1,050 15:43:52
or so periods of time they are available for 15:43:56
follow-up. 15:43:58

And what we do is we construct from the 15:44:02
records, from the follow-up survey of NHANES, the 15:44:08
period of time we know where they are and we know 15:44:12

whether they're in a nursing home or out of a 15:44:14
nursing home. 15:44:16

And then what we do is we stratify by 15:44:20
gender and by one-year age classes. So, for 15:44:28
example, we take all the 70 or all the persons who 15:44:30
are 70 years old during this ten-year period of 15:44:32
follow-up, and during their 70th year we calculate 15:44:36
the fraction of time that they were in a nursing 15:44:38
home, which is the number of days in a nursing home 15:44:44
divided by the total number of days they were in our 15:44:46
follow-up. That's an important point. 15:44:48

If we only knew that they -- where they 15:44:52
were for 180 days, and they were in a nursing home 15:44:56
for 90 of those days, that would be a ratio of a 15:44:58
half. And then we average those ratios to get a 15:45:04
fraction of time in the nursing home. 15:45:08

And when I said we stratify, we stratify 15:45:10
on gender, year, age, class and smoking, of course, 15:45:14

as well. 15:45:18
And then for each age and gender, we take 15:45:22
the fraction of time for the smokers minus the 15:45:28
fraction of time for the nonsmokers divided by the 15:45:30
fraction of time for the smokers, and that's the SAF 15:45:36
which we apply to the expenditures for persons in 15:45:40
that age category and in that gender. 15:45:42

Q. I take it that in your example people could be in 15:45:46
their 70th year in different calendar years? 15:45:50

A. Yes. 15:45:50

Q. And there are no factors controlled for here, other 15:46:02
than age, gender, smoking? 15:46:08

A. Yes, correct. 15:46:10

Q. There are other factors in NHANES, do you know? 15:46:16

A. I assume that there are. 15:46:18

Q. And in the case of lung cancer in the refined model 15:46:34
you explained to me why you didn't control for any 15:46:40
other factors. 15:46:40
I think it was, I'll paraphrase, but 15:46:40
because according to Dr. Samet the relationship 15:46:44
between smoking and those diseases was so strong. 15:46:48
What's the reason here? 15:46:50

A. In our discussions, it was, we thought, the most 15:47:00
important factor to admission, dominant factor to 15:47:06
admission, to a nursing home is age. And so we 15:47:12
stratified on that. And gender, we stratified on 15:47:18
that. 15:47:18
There was limited numbers of persons 15:47:24

available with the nursing home, you know, who 15:47:30
qualified for this follow-up study. 15:47:32
And I think because of the limited, I say 15:47:38
limited, because of the roughly a thousand persons 15:47:42

or so, I think people felt it was better to fit a 15:47:46
simpler model than a more complex one. 15:47:48
There were some -- well, those were the 15:47:54
main reasons. 15:47:54
Q. Dr. Miller told us, I think, that -- it was not 15:48:06
inconsistent with what you're saying, but he added 15:48:08
that you ran this many different ways and got 15:48:14
greatly variable results depending on the 15:48:18
specification, and that as a result you decided on a 15:48:24
very simple model. 15:48:26
Now, I don't warrant that's exactly what 15:48:28
he said and we can look at that. But does that ring 15:48:30
any bells as to other factors, other reasons? 15:48:32
A. I don't recall whether -- he tried one or two 15:48:40
approaches to this problem. I think it was perhaps 15:48:44
as I was coming into this and getting involved. 15:48:48
I don't recall whether the discussion was 15:48:52
that different variables that controlled for made -- 15:48:58
Q. He just said different specification got very 15:49:00
different results. 15:49:02
A. Like I say, I don't know specifically what it was 15:49:04
that led to the conclusion. When I got involved in 15:49:08
the discussions of the nursing home, it was, you 15:49:10

know, we don't have a lot of data here, we should 15:49:12
try to do something that's simple and direct, as 15:49:14

opposed to trying to build a complex model, a more 15:49:18
complex model like was done for the NMES data where 15:49:22
we had considerably more information. 15:49:30

Q. Now, at the beginning of the day, I think, or 15:49:38
earlier in the day we read together from page 4 of 15:49:56
your report where you said that -- I'm going to just 15:50:04
omit the word Blue Cross here from my reading -- 15:50:08
"The state paid substantially more nursing home 15:50:10
residence fees for smokers than for never-smokers." 15:50:14
Do you see that? 15:50:14

A. Yes. 15:50:14

Q. And, in fact, what you calculated was $238 million; 15:50:26
is that correct? 15:50:26

A. I don't remember the exact number. I'm sure it's 15:50:30
here someplace. 15:50:30

Q. Yeah, it is in paragraph 7, actually. 15:50:34

A. Where in paragraph 7, sir? 15:50:42

Q. Last sentence. 15:50:42

A. Two hundred thirty-eight, yes, point four million. 15:50:44

Q. So I take it that is the substantially more? 15:50:50

A. That's our estimate of it, yes. 15:50:52

Q. More than what? 15:50:54

A. We estimate that -- I'm sorry, maybe I was confused 15:51:02
in my last answer. We estimate that of the total 15:51:06
expenditures by Medicaid for maintenance of persons 15:51:12

in nursing homes during this period 1978 to 1996, 15:51:18
238.4 million of those expenditures are attributable 15:51:22
to smoking. 15:51:22

And that's 238 million out of the total 15:51:28
expenditures, which I don't have in my copy but 15:51:30
which are in the report. 15:51:30

Q. Well, I'm referring back to the first sentence in 15:51:36
paragraph 4 now where it says, "The state paid 15:51:38
substantially more nursing home residence fees for 15:51:42
nonsmokers than for never-smokers." Is it 238 15:51:48
million more? 15:51:48

A. No. 15:51:48

Q. How much more? 15:51:52

A. I don't know. 15:51:54

Q. You mean there is not a correspondence between that 15:51:58
sentence and the 238 million? 15:52:02

A. I don't know what the direct correspondence is. 15:52:14

Q. Do you know one way or another whether -- do you 15:52:20
know one way or another? 15:52:20

A. What? 15:52:22

Q. Why do you say that it's not 238 million more? 15:52:26

A. Because -- well, what 238 million is is the fraction 15:52:30
of the actual expenditures by the state, which based 15:52:36
upon our model built in NHANES, is attributable to 15:52:40
smoking. 15:52:42

I don't think, as I sit here, that that's 15:52:44
the same as saying that 238 million more was paid 15:52:50
for residence fees of smokers than for 15:52:56
never-smokers. 15:52:56

I don't think those two are equivalent 15:52:58
statements. 15:52:58

Q. Well, do you think that the state paid substantially 15:53:02
more nursing home residence fees for smokers than 15:53:04
for never-smoker? 15:53:04

A. That is the -- what it says in the expert report. 15:53:08
And I, at this moment, don't recall exactly what the 15:53:10
basis -- I think the basis of that sentence is the 15:53:14
results of the modeling, but the number is not 238 15:53:18
million. 15:53:18

Q. But if it's the result of the modeling, what result 15:53:26
of the modeling is it? What other result do we 15:53:30
have?

A. I don't know. 15:53:32

Q. What is the difference between the two -- between 15:53:48
the two statements, the last sentence of 7 and the 15:53:54
first sentence of 4? 15:53:54

A. Well, the difference is that the last sentence in 15:54:02
paragraph 7 says that we estimate that of the total 15:54:10
expenditures by the state for this period of time, 15:54:18
1976 -- excuse me, for this period of time that 15:54:24

we're considering, of the total expenditures, 238.4 15:54:28
million of them are attributable to smoking. 15:54:34

And I've described for you the logic by 15:54:36
which we have arrived at that number. The first 15:54:42
sentence in paragraph 4 says something about 15:54:46
residence fees paid for by the state and the amount 15:54:52
of fees paid for by smokers and never-smokers. 15:54:56

And those are not -- I don't think at this 15:55:00
point in time those are equivalent statements. And 15:55:02
I can't recall further what the basis for the 15:55:06
sentence at the top of the paragraph -- 15:55:08

Q. Your concern is not residence fees, is it? 15:55:14

A. No, my concern is not residence fees. 15:55:16

Q. Okay.

A. I may be missing something by the lateness of the 15:55:24
hour, but as I sit here I do not think that they're 15:55:26
equivalent. And I don't recall further the basis 15:55:32
for the first sentence in paragraph 4. 15:55:32

Q. Well, is it possible that the state did not pay more 15:55:36
nursing home fees for smokers than never-smokers? 15:55:38

A. I don't know. I don't know right now. 15:55:40

Q. So this statement may just be flat wrong? 15:55:42

A. I don't know. I don't think it's wrong, but I don't 15:55:44
know.

Q. Is it possible that the state did not pay more 15:55:56

nursing residence fees for smokers than for 15:55:58
never-smokers, and yet there are $238 million of 15:56:00
smoking attributable costs? 15:56:02

A. I don't know. 15:56:04

Q. I'm not asking whether that's true. I'm asking if that's possible? As a statistician, can that be? 15:56:06 15:56:08

A. I don't know. 15:56:10

Q. I take it that by saying you don't know that you don't rule it out. How could that be? 15:56:26 15:56:28

A. I'm not ruling it in or I'm not ruling it out. I don't know at this moment. 15:56:30 15:56:32

Q. Well, let's ask it another way. You say there's 238 million of smoking attributable nursing home maintenance care, correct? 15:56:42 15:56:44 15:56:46

A. Correct. 15:56:46

Q. And I take it that that means that if there hadn't been any smoking, that there would have been $238 million less for nursing home maintenance care? 15:56:52 15:56:54 15:56:58

A. No. 15:57:00

Q. So there's $238 million attributable to smoking, but there would not be $238 million less if there had not been smoking, correct? 15:57:10 15:57:16 15:57:18

MR. HAMLIN: Objection; asked and answered. 15:57:20 15:57:20

BY MR. SILFEN:

Q. Is that your position? 15:57:22

A. That's what I said, yes. Excuse me, would you repeat that question? 15:57:26 15:57:28

Q. You see; apparently, it wasn't asked and answered. 15:57:32

A. I think it was asked and answered, but I just want to make sure you didn't change the question because I didn't hear it exactly right. 15:57:36 15:57:38 15:57:40

(The requested portion read back.) 15:57:58

THE WITNESS: There are $238 million 15:57:58
attributable to smoking. That is not the same as 15:58:02
saying that there would have been $238 million less 15:58:04
had there been no smoking. 15:58:06

BY MR. SILFEN:

Q. Why? 15:58:06

A. Because I don't know what would have happened in the 15:58:08
world had there been no smoking. It would require 15:58:10
lots of speculation about things that might have 15:58:14
happened, which I'm not in a position to do. 15:58:16

Q. So your model does not tell us whether there would 15:58:20
have been more or less nursing home costs for the 15:58:26
state had there been no smoking, correct? 15:58:28

A. The model tells us the fraction of the actual 15:58:30
expenditures that did occur that are attributable to 15:58:34
smoking, which is not the same as the -- well, 15:58:42
that's what it tells us. 15:58:42

Q. For purposes of the record, I'd like an answer to 15:58:44
the question, and I'm not meaning to press you on 15:58:46
it. 15:58:48

But is it correct, then, that your model 15:58:50
does not tell us one way or another whether there 15:58:52
would have been more or less nursing home costs to 15:58:56
the state had there been no smoking? 15:58:58

MR. HAMLIN: Objection; asked and 15:59:00
answered. 15:59:00

THE WITNESS: Our model doesn't address 15:59:04
that question. 15:59:04

BY MR. SILFEN:

Q. Does your model tell us whether smokers or 15:59:08
nonsmokers in the period under study were more 15:59:12
likely to enter nursing homes in their lives? 15:59:14

A. It does not address that question specifically. 15:59:16

Q. Does the model tell us if smokers or nonsmokers were 15:59:28
likely to spend more time in nursing homes in their 15:59:32
lives? 15:59:32

A. It does not address that question specifically, 15:59:34
either. 15:59:34

Q. Does the state tell us -- does the model tell us if 15:59:42
the state spent more on smokers than on nonsmokers 15:59:46
for nursing home care? 15:59:48

A. It does not address that difference directly, 15:59:52

either. 15:59:52

Q. As far as you know, is it entirely possible that 16:00:16
smokers, in fact, spend less time in nursing homes? 16:00:20

A. Well, this model looks at the fraction of time at a 16:00:30
given age that smokers are in nursing homes relative 16:00:34
to nonsmokers. 16:00:36

And by virtue of the fact that this is a 16:00:38
positive $238 million means that when averaged over 16:00:44
all of the people for whom we had expenditures, that 16:00:50
there's at a given age, gender, there's an average 16:00:56
more time spent in nursing homes by, a larger 16:01:04
fraction of time by the smokers than by the 16:01:04

nonsmokers at a given age. 16:01:06

Q. But smokers are not -- you don't know one way or 16:01:10
another whether smokers are more likely to enter a 16:01:12
nursing home in their lives? 16:01:14

A. This calculation does not address that directly. 16:01:16

Q. Suppose the state said to you, "Dr. Zeger, we paid 16:01:34
you a bit of money here to do work for us, we want 16:01:38
to know will we have more or less nursing home costs 16:01:40
if there is no smoking," what would your answer be? 16:01:44

A. Well, they've not paid me any money to address that 16:01:46
question. And if they wanted me to address that 16:01:48
question, it would be a very expensive adventure. 16:01:52

I mean, I'd have to start and address a 16:01:54

different question. It's not the one I've addressed 16:01:56
here or we've addressed here. 16:02:00

Q. And how was it decided that you would not address 16:02:10
any of those questions that we have gone through 16:02:18
here? 16:02:18

A. We addressed the same question here that we 16:02:20
addressed in all the other parts of the analysis 16:02:24
that we've talked about today. 16:02:26

We said what would be the actual 16:02:28
expenditures for the state and what fraction of 16:02:30
those are attributable to smoking. 16:02:34

Q. When you compared the smoker with the nonsmoker on 16:02:40
an annual basis, you took the fractional part of a 16:02:46
year spent in a nursing home by a smoker and you 16:02:50

subtracted the fractional part of a year spent in 16:02:54
the nursing home by a nonsmoker, correct? 16:02:56

A. We took all of the smokers and all of the nonsmokers 16:02:58
in a given stratum, yes. 16:03:00

Q. I didn't mean to make it personal. 16:03:02

A. Yeah, compared the difference in the fraction of 16:03:06
time they were in the nursing home. 16:03:08

Q. So you compared, for the stratum you were looking 16:03:16
at, you compared a smoker with a never-smoker, 16:03:18
correct? 16:03:18

A. No. 16:03:22

Q. You compared a smoker time in a nursing home for the 16:03:28
given age group to the nonsmoker or the never-smoker 16:03:34
time in a nursing home for a given age group, 16:03:38
correct? 16:03:38

A. For a given age and gender we took all of the 16:03:40
smokers and all of the nonsmokers and we took the 16:03:44
average fraction of days for the smokers and 16:03:48
compared that to the average fraction of days for 16:03:52
the nonsmokers, never-smokers, in that age and 16:03:54
gender category. 16:03:56

Q. And you assumed that the difference between those 16:04:00
two was an extra attributable to smoking, right? 16:04:08

A. We assumed that the difference between those two, as 16:04:12
a ratio to the bigger, is the fraction of additional 16:04:18
time that's attributable to smoking. 16:04:20

Q. So what you've said is that -- you've calculated, I 16:04:30
take it, on an annual basis, the cost that the state 16:04:38

http://legacy.library.ucsf.edu/tid/iqq07a00/pdf

would have had or would have saved if smokers had 16:04:44
had the same nursing home stay as nonsmokers on 16:04:50
average for that year? 16:04:52
A. No. 16:04:52
Q. Oh, well, I -- let's take one year. Let's say there 16:05:00
was only one year in the study, people 70 years 16:05:02
old. Okay? 16:05:04
A. Okay. 16:05:04

Q. For those people 70 years old, let's say the average 16:05:08
nursing home stay was ten days for smokers and five 16:05:12
days for nonsmokers. Okay? 16:05:16
A. Okay. 16:05:16
Q. That's the average over all of the people in that 16:05:20
stratum. So the excess days that you would say were 16:05:24
attributable to smoking is five, correct? 16:05:26
A. We calculated it in a fraction, an excess fraction 16:05:32
of days. 16:05:32
Q. I understand. I could put in the denominator and 16:05:36
the denominator would be 10. 16:05:36
A. The denominator would be 10 in a year? 16:05:40
Q. Well, what you would -- what the denominator would 16:05:42
be is the days that the smoker spent, the average 16:05:48
days that the smoker spent in the nursing home? 16:05:50
A. That's not what we did. I mean, I don't mean to be 16:05:54
argumentative, but that's not what we did. 16:05:56
Q. That's fine, then tell me what you did. 16:05:58
A. What we did is we took for each person the fraction 16:06:02

of time they were in. So if the person was there in 16:06:06
our study for a whole year, and they were in for 180 16:06:08
days, their fraction was a half, I mean, roughly a 16:06:12
half. 16:06:12

If somebody was only in our study for two 16:06:14
days and they were in the nursing home for one day, 16:06:16

that was also a half. So you can't just take the 16:06:18
number of days. You have to look at the 16:06:20
denominator, as well, which is the total days they 16:06:24
were at risk of being in a nursing home. 16:06:26

And we calculated the average fraction for 16:06:28
the smokers and compared that to the average 16:06:30
fraction for the nonsmokers, which is different from 16:06:34
taking the numbers of days for the smokers and 16:06:38
subtracting the number of days for nonsmokers. 16:06:40

Q. And the difference between those two fractions you 16:06:44
said was the excess nursing home time due to 16:06:48
smoking? 16:06:48

A. We said it was the fraction, that thing divided by 16:07:00
the average fraction of time for the smokers was the 16:07:02
smoking attributable fraction that we applied to the 16:07:06
expenditures for nursing home stays. 16:07:08

Q. And so for each year you were charging smoking with 16:07:18
responsibility for any difference between the 16:07:26
fraction of time spent by a smoker, the smokers 16:07:32
population wide, and nonsmokers? 16:07:34

A. Correct. 16:07:34

Q. And why is that extra, that difference which we've 16:08:32

just identified, not the difference for that year 16:08:42
that is attributable to people smoking? 16:08:48

A. I don't understand that question. 16:08:54

MR. HAMLIN: Objection. 16:08:58

BY MR. SILFEN:

Q. Is that difference a difference that you say is 16:09:00
attributable to smoking? 16:09:00

A. Yes, and it's for a year of age. 16:09:04

Q. Is that something different than saying that it is 16:09:06
attributable to the fact that people smoke? 16:09:08

A. No. Similar; different words. 16:09:14

Q. And then is it not also true that that difference 16:09:24
would disappear if people did not smoke? 16:09:28

A. That is not necessarily true. 16:09:32

Q. So you say that is a difference that is attributable 16:09:34
to people, to the fact that people smoked, but that 16:09:38
difference would not disappear if people did not 16:09:42
smoke? 16:09:42

A. That's correct. 16:09:44

Q. How can that be? 16:09:46

A. Well, there are lots of factors that might change if 16:09:52
there was no smoking in the world. And in order to 16:09:56
make this comparison, I have to have a certain 16:09:58
population of people, and I have to have smokers and 16:10:02
nonsmokers, and you're talking about a different 16:10:04
world. 16:10:06

I don't know that the data would look the 16:10:08

way it does if we had a different world without 16:10:10

smoking. So I can't tell you that they're the 16:10:14
same. 16:10:14

Q. And why didn't you look at those other factors that 16:10:24
you described that would have changed? 16:10:28

A. Well, because I can only speculate about what might 16:10:32
have changed. And it wasn't relevant to the 16:10:34
question that I addressed in this report. 16:10:36

Q. Well, which is it or both? Is it that it's not 16:10:40
relevant or that it was speculative? 16:10:42

A. It's both. It's both. The question -- I was able 16:10:48
to address the question, we were able to address the 16:10:50
question that was asked, which had to do with 16:10:52
attributing the actual expenditures without having 16:10:56
to speculate about how things might have been in 16:11:00
another world, and that's why we approached the 16:11:04
problem the way we did. 16:11:10

MR. SILFEN: Why don't we take five 16:11:14
minutes. 16:11:16

(A break was taken.) 16:11:18

THE VIDEOGRAPHER: We're back on the video 16:16:32
record. This is the fourth tape of the video 16:16:34
deposition of Scott Zeger. The time is now 4:16 16:16:38
p.m. 16:16:40

BY MR. SILFEN:

Q. Back on the record. Having given you an early 16:16:44

warning, you do know what a life table is? 16:16:46

A. Yes. 16:16:46

Q. Are life tables speculative? 16:16:48

A. They can be. 16:16:50

Q. Are they any more speculative than the data in 16:16:54
NHANES or in NMES? 16:16:58

A. A life table can be filled with speculative numbers, 16:17:02
or it can be filled with numbers estimated 16:17:06
accurately from data. 16:17:06

Q. Are there reliable life tables for smokers and 16:17:10
nonsmokers? Are you familiar with ACS million 16:17:14
person study, for instance? 16:17:16

A. Yes, I am. 16:17:16

Q. Do you believe that there are reliable life tables 16:17:18
for smokers and nonsmokers?

A. I don't know the extent to which there are reliable 16:17:22
life table studies. I'm somewhat familiar with that 16:17:24
study, but I don't know about the life tables. 16:17:26

Q. I was just picking a study out of the air. But is 16:17:28
it your position that there are not reliable life 16:17:32
tables for smokers and nonsmokers? 16:17:34

A. It's not my position that there are or that there 16:17:36
aren't. 16:17:36

Q. You don't know one way or the other? 16:17:38

A. I don't know about the life table data on smokers or 16:17:42

nonsmokers. 16:17:42

Q. If there are reliable life table information for smokers and nonsmokers, say as reliable as NMES and NHANES, then why would the life expectancy difference between smokers and nonsmokers be speculative? 16:17:46 16:17:50 16:17:54 16:17:58 16:17:58

MR. HAMLIN: Objection to foundation. 16:18:00

THE WITNESS: I didn't say that the, conditioned on the premise, that there are accurate life tables for smokers and nonsmokers. I did not say that the life tables had to be speculative or using the life tables had to be speculative. 16:18:02 16:18:10 16:18:12 16:18:18 16:18:22

I used the word speculative in response to a particular question you asked me previously. 16:18:24 16:18:26

BY MR. SILFEN:

Q. Well, all right. Let's assume that if people didn't smoke, one of the things that would change would be their life expectancy. 16:18:28 16:18:30 16:18:34

Do you believe that to be true? 16:18:34

A. I do believe that, that's true, yes. 16:18:36

Q. Do you think the change in their life expectancy is speculative? 16:18:38 16:18:40

A. What? 16:18:42

Q. If they didn't smoke. 16:18:42

A. If they didn't smoke, I would -- I mean, from what 16:18:48

I've read, and I've not studied this topic, but from what I've read I would expect on average people who don't smoke live longer than people who do smoke. 16:18:52 16:18:54 16:18:58

Q. Do you have any reason to believe that the data 16:19:00
indicating the life expectancy difference between 16:19:04
smokers and nonsmokers is any more or less reliable 16:19:08
than the data indicating the disease incidence 16:19:16
difference between smokers and nonsmokers? 16:19:16

MR. HAMLIN: Objection to form and 16:19:16
foundation. 16:19:18

THE WITNESS: I'm sorry, you'll have to 16:19:22
ask it again. 16:19:22

BY MR. SILFEN:

Q. Let me ask it in two parts. You have used here NMES 16:19:30
to derive a comparison of smoker and nonsmoker 16:19:36
disease incidence experience, correct? 16:19:44

A. No. 16:19:44

Q. All right. You've used NMES to compare smoker and 16:19:48
nonsmoker rates of disease, correct? 16:19:52

A. For the prevalence rates, correct. 16:19:54

Q. Do you believe that the studies comparing smoker and 16:20:02
nonsmoker expected life span are any less reliable 16:20:06
than the data you used in NMES to compare smoker to 16:20:10
nonsmoker? 16:20:10

A. I have no basis to know one way or the other. 16:20:12

Q. Let's assume that that data is as reliable, then why 16:20:14
would it be speculative when we compare smokers to 16:20:18
nonsmokers to also hypothesize what the difference 16:20:20
would be in their life-span if smokers didn't 16:20:24
smoke? 16:20:24

MR. HAMLIN: Objection to form, 16:20:26
foundation, assuming facts not in evidence. 16:20:28
THE WITNESS: I'm not sure I understand 16:20:30
the relevance of the question to what -- 16:20:30

BY MR. SILFEN:

Q. Well, then I will tell you the relevance. You said 16:20:34
that the reason you didn't consider another world, a 16:20:38
different world, is because it would be 16:20:40
speculative. That was one of your reasons, right? 16:20:42

A. I said there were two reasons why we did what we 16:20:46
did. 16:20:46

Q. I'm addressing one of them now. You said one of 16:20:48
them was that it was speculative, right? 16:20:50

A. I said it would be more speculative, yes. 16:20:52

Q. And I'm addressing that now. And my first point was 16:20:56
that you used NMES to compare smokers and 16:20:58
nonsmokers, correct? That's the point I just made. 16:21:04
Okay. 16:21:04
And then I asked you whether the life 16:21:06
table data on smokers and nonsmokers was any more 16:21:08

speculative than the data in NMES you relied on, and 16:21:12
you said you didn't know? 16:21:12

A. That's correct, that's what I said. 16:21:14

Q. So do you have any reason to believe that it would 16:21:18
be more speculative to hypothesize what would happen 16:21:24
to life-span if smokers didn't smoke than it would 16:21:28
what would happen to disease rates? 16:21:30

MR. HAMLIN: Objection to form and 16:21:32

foundation. 16:21:32

THE WITNESS: I mean, the question is, 16:21:34
given that there's no difference in the precision 16:21:36
between the NMES data and this life table data, is 16:21:40
there any difference in the precision, is there any 16:21:42
more speculation? 16:21:44

BY MR. SILFEN:

Q. I understand, it's a tautological question. 16:21:48

A. You know, what's the purpose of -- if A is true, 16:21:54
then A is true. 16:21:54

Q. You told me that age is the principal determinant of 16:21:58
nursing home entry, didn't you, you said that? 16:21:58

A. I did not say that, no. 16:21:58

Q. You said age is a main determinant, right? 16:22:02

A. I did not say that, no. I said age is a main 16:22:04
determinant of the probability somebody is in a 16:22:06
nursing home. I did not say -- I was talking about 16:22:08

prevalence not incidence. 16:22:10

Q. Well, okay, I'll take it. Well, we can -- do you 16:22:18
want to look for what you said? Why don't we look 16:22:20
for what he said. 16:22:20

(The requested portion read back.) 16:22:58

MR. SILFEN: "In our discussions, it was,
we thought, the most important factor to admission,
dominant factor to admission, to a nursing home, is
age."

BY MR. SILFEN:

Q. Now, if the most important factor is age and if the 16:23:06
age difference, expectancy difference between 16:23:10
smokers and nonsmokers, is not speculative, why 16:23:14
wouldn't we take that into account? 16:23:14

A. I'm sorry, repeat the question. 16:23:22

MR. HAMLIN: Objection; form. 16:23:42

THE WITNESS: For what purpose? 16:23:42

BY MR. SILFEN:

Q. You said that you didn't answer the question of what 16:23:50
would have happened had people not smoked, in part, 16:23:54
because it was speculative, because of the other 16:23:58
factors that might be involved, correct? 16:24:00

A. What I said was that to address the question that we 16:24:06
were asked, what fraction of actual expenditures are 16:24:10
attributable to smoking, it was not necessary to 16:24:14

look at these. Well, I didn't say -- what I -- 16:24:22

Q. Dr. Zeger, you also said it was speculative, and 16:24:26
that's the part I'm addressing. I'm not speaking to 16:24:28
the nature of your task. I'm speaking to your 16:24:30
statement that it was speculative. 16:24:34

A. What I'm trying to add and say to you now is that 16:24:36
it's -- it was more speculative to have to conjure 16:24:42
up what might have been than it was to use the data 16:24:44
which directly addressed the question that we had at 16:24:48
hand. 16:24:48

Q. The data you used compared smokers and nonsmokers 16:24:52
for their rate -- their different rate of disease, 16:24:58
correct? 16:24:58

A. For their rate of being in a nursing home. 16:25:02
Q. Okay. Their rate of being in a nursing home? 16:25:04
A. Correct. 16:25:04
Q. And why is it more speculative to address the 16:25:10
additional factor of their different life expectancy 16:25:12
if they didn't smoke? 16:25:14
MR. HAMLIN: Objection to form and 16:25:16
foundation. 16:25:16
THE WITNESS: The point I made, the 16:25:18
comment I made about speculative has to do with 16:25:22
trying to figure out how the world might have been 16:25:24
if people didn't smoke. 16:25:26

One of the aspects of that world might 16:25:28
have been their life expectancy and other aspects 16:25:30
might have been other things. 16:25:32
BY MR. SILFEN:
Q. But the point I'm making is that it is not at all 16:25:34
speculative or at least certainly no more 16:25:36
speculative than the data that you use? 16:25:42
A. I disagree with that. 16:25:42
Q. Really? Do you think the data comparing smoker and 16:25:46
nonsmoker disease risk is more or nursing home risk 16:25:50
is more reliable than the data on different -- on 16:25:52
life expectancies for smokers and nonsmokers? We 16:25:56
just went through. You said you didn't know. 16:25:56
MR. HAMLIN: Objection; form. I also 16:25:58
object on the grounds the question is 16:26:00

argumentative. 16:26:00

BY MR. SILFEN:

Q. I agree with you it's been asked and answered, he said he didn't know. Now, do you or do you not know? 16:26:02 16:26:06 16:26:06

A. What? 16:26:06

Q. Is the data on disease rates, difference between smokers and nonsmokers, any more reliable that the life expectancy rates? 16:26:10 16:26:14 16:26:18

MR. HAMLIN: Same objections. 16:26:18

THE WITNESS: I answered I don't know and I don't know. 16:26:20 16:26:20

BY MR. SILFEN:

Q. Do you believe that the NHANES smoker population and nonsmoker population that you examined had a different age distribution? 16:26:36 16:26:40 16:26:44

A. I don't know. 16:26:46

Q. Why wouldn't it be a simple solution to this problem to simply leave them with whatever age distribution they had? In other words, not age standardized, and compare their nursing home days? 16:27:00 16:27:02 16:27:10 16:27:12

A. Because that fails to control for the most important factor for being in a nursing home as opposed to admission to a nursing home for prevalence of nursing home days, which is age. 16:27:16 16:27:18 16:27:22 16:27:24

Q. Do you believe that if people stopped smoking that age difference between them, if there is one, would persist? 16:27:26 16:27:28 16:27:28

A. I don't know. 16:27:30

Q. If any age difference between smokers and nonsmokers would go away if they stopped smoking, why would you age adjust? What sense does that make? 16:27:34 16:27:38 16:27:40

MR. HAMLIN: Objection to form and foundation, also repetitive. 16:27:42 16:27:44

THE WITNESS: Because it allows me to

address the question which was asked. 16:27:46 16:27:48

BY MR. SILFEN:

Q. I see. 16:27:50

A. Which is what fraction of the dollars actually expended for nursing home are attributable to smoking. 16:27:54 16:27:58 16:27:58

Q. Suppose that their age expectancy is a function of their smoking. Is it then correct or even appropriate to age adjust? 16:28:04 16:28:12 16:28:14

A. If the question that you're asking is the question that we addressed, yes. 16:28:18 16:28:18

Q. So basically you were doing what you were told and it has nothing to do with speculation or data or anything else; is that right? 16:28:28 16:28:30 16:28:32

A. I was choosing the best statistical methods to address the question which was asked, which was what is the smoking attributable expenditures for smoking? 16:28:36 16:28:38 16:28:40 16:28:40

Q. Now, one of our experts approached the problem differently, took the smokers and nonsmokers as he 16:29:34 16:29:42

found them, and for the ten-year period counted up 16:29:50
nursing home days for the smokers and nursing home 16:29:54
days for the nonsmokers. 16:30:00

And in each case, the denominator was the 16:30:04
people starting out, so he basically got an average 16:30:08

nursing home days for the ten-year period for all 16:30:14
smokers, whether they went in or not, and all 16:30:18
nonsmokers, whether they went in or not. 16:30:20

Are you with me? I'm not asking you to 16:30:22
approve, I'm just asking if you understand what I'm 16:30:24
saying.

A. I'm not 100 percent sure what he did, no. 16:30:28

Q. He just took, he looked at the whole ten-year 16:30:30
period, took the smokers, counted up their nursing 16:30:32
home days, took the nonsmokers, counted up their 16:30:36
nursing home days, and he averaged those over all 16:30:40
the smokers who started and all the nonsmokers who 16:30:44
started. 16:30:44

A. Right. 16:30:44

Q. So what it was, in one I believe that the effect is 16:30:50
to measure for the ten-year period both probability 16:30:56
of entry and duration of stay and takes into account 16:31:02
smoking prevalence, as well? Doesn't it do all 16:31:04
those things? 16:31:06

A. I don't know. 16:31:06

Q. Okay. Now, his finding was that nonsmokers were 16:31:16
much more likely to enter nursing homes and stayed 16:31:22
longer when they entered. Do you have any reason to 16:31:24

doubt that result? 16:31:26

A. I don't have any reason to believe it nor to doubt 16:31:30

it. I don't know. 16:31:30

Q. Okay. What question does that analysis answer? 16:31:34

MR. HAMLIN: Objection to form. 16:31:38

THE WITNESS: That analysis would address 16:31:44 the average numbers of days that a nonsmoker was in 16:31:56 a nursing home minus the average numbers of days a 16:32:02 smoker was, without paying attention to what's the 16:32:06 critical factor in life table or other analysis, 16:32:10 which was the time that a person was at risk to be 16:32:12 there. 16:32:14

BY MR. SILFEN:

Q. Well -- 16:32:16

A. I think. 16:32:20

Q. Well, I don't understand that, but we'll come back 16:32:24 to that. 16:32:24

Let's mark as the next exhibit the chart 16:32:40 done by Tim Wyant that we referred to earlier in 16:32:48 connection with a statement in your report, 16:32:56 paragraph 4 on page 4. It's the second sentence in 16:33:14 that paragraph 4, "smokers entering nursing homes," do you see that?

A. Yes.

Q. And you recall we talked about that earlier? 16:33:18

A. I do recall indicating that I thought Tim had done 16:33:20 some calculations related to that. 16:33:22

Q. And you said you'd looked at it -- 16:33:24

A. Briefly. 16:33:26

(Defendants' Exhibit 2403 marked for 16:33:44
identification by the reporter.)

BY MR. SILFEN:

Q. I will give you Exhibit 2403 and ask you if that is 16:33:50
the chart that you remember looking at earlier? 16:33:56

A. I believe so. 16:34:12

Q. And what do you understand the chart to represent? 16:34:16

A. I believe this represents the diagnosis of persons 16:34:28
entering -- upon entrance to a nursing home 16:34:32
comparing nonsmokers to smokers. 16:34:38

I believe it gives the numbers of persons 16:34:40
with different smoking attributable diseases. And I 16:34:44
believe it's upon entrance to nursing home, but I'm 16:34:48
not 100 percent certain. 16:34:48

Q. But it was over the entire ten years of the 16:35:00
follow-up, I believe, I think from what I recall. 16:35:06
Let me back up. 16:35:10

The study I described as having been done 16:35:14
by a defense expert, is that a longitudinal study of 16:35:22
nursing homes? 16:35:22

A. I think I would call that a longitudinal study. 16:35:26

Q. And also an incidence study? 16:35:28

A. No, I don't think I would call it an incidence 16:35:38

study. 16:35:40

Q. Why not? 16:35:42

A. Well, you know, you have to be more specific about 16:35:48
what he did. 16:35:48

Q. That's fair. That's fair, why don't we wait. Is 16:35:58
this a longitudinal study that Dr. Wyant did? 16:36:02

A. Well, the data that we used in our calculations were 16:36:08
longitudinal data, yes. 16:36:10

Q. Is this an incidence study that Dr. Wyant did? 16:36:12

A. No. 16:36:14

Q. At the bottom you see for nonsmokers the totals, 16:36:32
persons with any tobacco related disease and persons 16:36:36
with no tobacco related disease, do you see that? 16:36:40

A. I see those two rows, yes. 16:36:42

Q. And the total there would be 415 persons, 313 plus 16:36:50
102? 16:36:50

A. For nonsmoking? 16:36:54

Q. Yes. 16:36:54

A. 415 would be the sum of any tobacco-related diseases 16:36:58
and no tobacco-related diseases. 16:37:00

Q. And for smokers how many people do we have entering 16:37:06
nursing homes? 16:37:06

A. Let's see, by my calculations 259, the sum of 213 16:37:18
and 46. 16:37:20

Q. Now, earlier today we talked about the data that you 16:37:26

had reported on the prevalence of smoking in 16:37:30

Minnesota. Do you remember that? 16:37:30

A. I don't recall that discussion, no. 16:37:38

Q. Well, we looked together at chart 1, Table 1, on 16:37:44
page 2 of your report. 16:37:46

A. Yes, this is for persons 19 and above for the period 16:37:56
1984 to '94. 16:37:58

Q. Okay. And I think what you told me that for persons 16:38:00
19 and above during this period at least 50 percent 16:38:06
were ever-smokers, right? 16:38:08

A. Yes, in these two categories for persons 19 and 16:38:14
above, more than 50 percent were current or former 16:38:18
smokers. 16:38:18

Q. Let's assume that that 50/50 split persists right up 16:38:24
through the age groups. I don't know that it does, 16:38:28
but let's assume that that is correct. 16:38:30

Would Dr. Wyant's paper then tell us that 16:38:38
nonsmokers entered nursing homes during this NHANES 16:38:44
period at, what, one and a half times the rate of 16:38:50
nonsmokers, of smokers? 16:38:52

A. I don't know. 16:38:58

Q. Well, you don't know because we haven't done the 16:39:00
math or -- we can do the math. 16:39:02

A. I don't know. 16:39:04

Q. Well, I'll do the math. No, I won't. I think it's 16:39:22

1.6. Let's assume we had a calculator and we 16:39:34
divided 415 by 259 and we got 1.6. 16:39:40

Would Dr. Wyant's study then tell us that 16:39:44
over the NHANES period that nonsmokers entered 16:39:50

nursing homes at about a 60 percent greater rate 16:39:52
than smokers? 16:39:52

A. I don't know that that's what I would say. 16:40:00

Q. Why? What is it that we don't know? 16:40:02

A. Well, I don't know specifically what assumptions 16:40:06
you're making or what calculation that you're 16:40:08
making. 16:40:08

You've calculated the ratio 415 to 259 and 16:40:14
you've said -- I don't know what that means. 16:40:16

Q. Well, I'd like to understand this, so I didn't mean 16:40:20
to make assumptions. 16:40:22

My assumptions are this, that Dr. Wyant 16:40:26
has calculated that during the NHANES follow-up 16:40:32
period 415 nonsmokers entered nursing homes and 259 16:40:38
smokers entered nursing homes. 16:40:42

I don't think that's an assumption. 16:40:44
That's what the data reflects, right? 16:40:46

A. That's my understanding, yes. 16:40:48

Q. And I have assumed that the 50/50 smoker/nonsmoker 16:40:56
split for Minnesotans, which we see in your report, 16:41:00
persists through all age groups. That is the 16:41:04

assumption that I'm making. Okay? 16:41:06

A. I don't know whether that assumption is even close 16:41:08
to -- 16:41:10

Q. Let's assume that it is. 16:41:12

A. I don't know. 16:41:14

Q. Let's assume it's right. I want to understand how 16:41:16

we do these calculations. 16:41:18
Would this study that Dr. Wyant did then 16:41:24
demonstrate that nonsmokers entered nursing homes 16:41:32
at -- well, at least a 50 percent greater rate than 16:41:36
smokers? 16:41:38
MR. HAMLIN: Objection; asked and 16:41:40
answered. 16:41:40
THE WITNESS: I wouldn't -- I mean, I 16:41:54
don't know what it would -- why that ratio would be 16:41:58
relevant. 16:41:58
BY MR. SILFEN:
Q. All right. Let's just say at a greater rate. 16:42:04
A. I can't conclude that. I don't know that that's the 16:42:06
case. 16:42:06
Q. And is the only reason because you don't know the 16:42:08
prevalence in the community? 16:42:10
A. That's one reason. 16:42:10
Q. What's the other reason? 16:42:12
A. I don't know who these people are. I don't know 16:42:18

their ages. I don't know anything else about them. 16:42:18
I mean, that data may be available in NHANES and 16:42:24
might well be taken into account. 16:42:24
Q. I'm not asking you anything about age. I'm asking 16:42:24
whether the smokers, the nonsmokers in NHANES, 16:42:26
entered nursing homes at a greater rate than the 16:42:30
nonsmokers? 16:42:30
A. Well, if you're asking me whether 415 smokers -- 16:42:36
nonsmokers entered and only 259 smokers entered, 16:42:42

that is to say if more nonsmokers entered, that's 16:42:46
clearly the case in this data. 415 is greater than 16:42:50
259; more nonsmokers entered than smokers. 16:42:58

If you're asking me to reach some 16:43:00
scientific conclusion based upon that fact, I'm not 16:43:04
willing to do so. 16:43:04

Q. Actually, I'm just trying to understand -- now, on 16:43:30
page 4 of your report, you said that smokers 16:43:36
entering nursing homes during this period were far 16:43:38
more likely, do you see that, than never-smokers to 16:43:42
be suffering from lung cancer and chronic 16:43:46
obstructive pulmonary disease, do you see that? 16:43:48

A. I do see that, yes. 16:43:50

Q. And do you see that the basis of that statement is, 16:43:52
in fact, these reports, this chart? 16:43:54

A. Let's see, we're talking about lung cancer for which 16:44:04

there were 15 persons who had lung cancer who were 16:44:12
smokers and no persons who had lung cancer were 16:44:16
nonsmokers, and then for COPD there were 109 16:44:18
smokers -- sorry, 109 COPD cases among the smokers 16:44:26
and 51 among the nonsmokers. 16:44:28

Q. And how many nursing home cases were there among the 16:44:30
nonsmokers, 415? 16:44:36

A. 415 total, yes. 16:44:38

Q. And how many nursing home cases were there among the 16:44:40
smokers, 259? 16:44:44

A. Right. 16:44:44

Q. So are we then -- have we then reason to conclude 16:44:48
that during this period nonsmokers were far more 16:44:54
likely than never-smokers to enter nursing homes? 16:44:56

A. No, because -- no. 16:45:04

Q. Why not? 16:45:04

A. Because I believe the basis for the sentence is 16:45:08
reflected in the fact that among the persons who 16:45:18
enter here, far more likely -- yes, among the 16:45:26
persons entering here, 12 percent had COPD among the 16:45:32
nonsmokers, as opposed to 42 percent among the -- 16:45:34

Q. This is just another way of saying we need the 16:45:38
prevalence number. But if we had the prevalence 16:45:40
number and the prevalence number is indeed 50/50, 16:45:42
wouldn't it be true that you could just as reliably 16:45:48

say that nonsmokers are far more likely to enter 16:45:52
nursing homes than smokers, as you could make 16:45:54
conclusions about lung cancer? 16:45:56

MR. HAMLIN: Objection; asked and 16:45:58
answered. 16:45:58

THE WITNESS: No. 16:45:58

BY MR. SILFEN:

Q. Why not? 16:45:58

A. Because my understanding of the sentence in 16:46:02
paragraph 4 is referring to the fact that among 16:46:08
those persons entering a nursing home, 259 smokers 16:46:12
and 415 nonsmokers, 42 percent of the smokers had 16:46:20
COPD, while only 12 percent of the nonsmokers did. 16:46:24
And 6 percent of the smokers had lung cancer, while 16:46:26

none of the nonsmokers did. 16:46:28
Q. Yeah, but if we assume that -- 16:46:30
A. I believe that's the point here. 16:46:32
Q. If we assume there's 1,000 nonsmokers in the 16:46:36
community and 1,000 smokers, we could just put them 16:46:38
over a denominator and get the same result? 16:46:42
A. No. No, that's not true. 16:46:44
Q. Why? 16:46:44
A. By taking that logic, which I've just used, to 16:46:46
justify this second sentence in paragraph 4, the 16:46:50
comparison is that 100 percent of the smokers went 16:46:54

into nursing homes, entered nursing homes, as 16:46:56
compared to 100 percent of the nonsmokers, which is 16:47:00
the same. 16:47:00
So you're asking me whether it's the same 16:47:02
thing, and I'm telling you that it's not. 16:47:04
Q. I'll try once more. I want you to assume that the 16:47:12
smoker/nonsmoker prevalence split in the community 16:47:16
is 50/50 throughout all ages. 16:47:20
Can you assume that, just a hypothetical? 16:47:22
You must work with them all the time? 16:47:24
A. I avoid them like the plague. 16:47:26
Q. If you will not accept that, tell me now because I 16:47:30
don't want to do -- will you accept that 16:47:32
hypothetical? 16:47:32
A. No. 16:47:32
Q. Why? 16:47:34

A. Because I don't have any evidence for which to base 16:47:36
that it's true. I mean, you're asking me to 16:47:38
speculate about something I don't know about and I 16:47:40
feel uncomfortable doing that. 16:47:42

MR. HAMLIN: Let him finish his answer. 16:47:42

THE WITNESS: You're asking me to 16:47:44
speculate about something I don't know one way or 16:47:48
the other, and that makes me uncomfortable as a 16:47:50
person trying to give you accurate answers. 16:47:52

BY MR. SILFEN:

Q. No, actually, you know, that really isn't a fair 16:47:54
answer. 16:47:54

I'm asking you to accept a predicate so 16:47:58
that we can do an analysis of data we do have. 16:48:02
Because I want to understand how the data works. 16:48:04

So I'm saying a predicate. I'm asking you 16:48:06
to accept as a predicate a 50/50 split, and then I'm 16:48:12
saying, all right, then how do we interpret it? 16:48:16
Now, you will not do that? 16:48:16

MR. HAMLIN: Objection; foundation. 16:48:16

MR. SILFEN: I understand.

MR. HAMLIN: There's no factual foundation 16:48:20
for your predicate. I mean, if you were going to 16:48:22
present him with some data -- 16:48:24

MR. SILFEN: I have presented his own 16:48:26
results that say smokers -- 16:48:28

MR. HAMLIN: No, no, present him with data 16:48:30
about prevalence, Tom, then there might be some 16:48:32

basis for the hypothetical. Right now you're asking 16:48:34
him to speculate. 16:48:34

BY MR. SILFEN:

Q. On page 2 of your report, it says that the 16:48:38
smoker/nonsmoker prevalence split for Minnesotans 16:48:42
age 19 and over is over 50 percent. 16:48:44

All I'm asking you to do is assume that 16:48:46
that continues through the years. 16:48:48

MR. HAMLIN: That chart does not deal with 16:48:50
nursing homes, and you know that. 16:48:52

MR. SILFEN: But that's not the issue. 16:48:54
It's not how many people are in nursing homes, it's 16:48:58
what the split is in the community. It's where they 16:49:00
draw from, Tom. 16:49:02

MR. HAMLIN: You're asking him a 16:49:04
hypothetical about nursing homes.

MR. SILFEN: But, Tom, you were wrong --

BY MR. SILFEN:

Q. Isn't it true that the critical denominator is 16:49:06
smokers and nonsmokers? Let's settle this. The 16:49:10
denominator that you would want to know, the 16:49:12
prevalence, would be the prevalence of smokers and 16:49:14
nonsmokers in the community, right? 16:49:16

A. For what purpose? 16:49:18

Q. To determine whether smokers and nonsmokers enter 16:49:20
nursing homes at different rates? 16:49:22

A. I would not ask the question to smokers and 16:49:26

nonsmokers entering nursing homes at different rates 16:49:30
without stratifying on the key known factors. 16:49:36

Q. Okay. Let's do this -- 16:49:36

A. The most important of which is probably age. 16:49:38

Q. You have said -- tell me what prevalence data you 16:49:48
need in order to complete this analysis. 16:49:52
Your report says Minnesotans 19 and over 16:49:54
and it gives a prevalence data. Tell me what age 16:49:58
group. Do I need smoking prevalence for Minnesotans 16:50:00
55 and over and then you'd be able to do it? 16:50:04

A. To do what? 16:50:04

Q. To tell me whether smokers or nonsmokers during the 16:50:08
period of NHANES entered nursing homes at a greater 16:50:12
rate. 16:50:12

A. I'm saying that I wouldn't ask that question. 16:50:16

Q. I'm asking you the question, and I want to know what 16:50:18
additional data you need. 16:50:20

A. And I don't know what additional data I need because 16:50:24
you're asking me to do something here on the fly. 16:50:26
And when you do things on the fly like that, you 16:50:30
make mistakes, and I don't want to make a mistake 16:50:32
that I'll regret. 16:50:32

Q. Okay. I wouldn't want you to make a mistake that 16:50:38
you would regret. 16:50:38

MR. HAMLIN: Can we go off for just a 16:50:42
minute?

MR. SILFEN: No.

MR. HAMLIN: We can't go off?

MR. SILFEN: Do you have another luncheon 16:50:42

date?

MR. HAMLIN: No. Off the record. 16:50:46

(Off the record.) 16:51:08

THE VIDEOGRAPHER: This concludes the 16:51:10
testimony of Scott Zeger. The time is now 4:51 16:51:14
p.m. 16:51:14

(The deposition was adjourned for the day.) 16:51:14

STATE OF MINNESOTA)
)
COUNTY OF HENNEPIN)

BE IT KNOWN THAT I, JENNIFER S. SATI, took the DEPOSITION OF SCOTT ZEGER, Ph.D., VOLUME I;

THAT, I was then and there a Notary Public in and for the County of Hennepin, State of Minnesota;

THAT, I exercised the power of that office in taking said deposition;

THAT, by virtue thereof I was then and there authorized to administer an oath;

THAT, said witness, before testifying, was duly sworn to testify to the truth, the whole truth, and nothing but the truth, relative to this action;

THAT, said witness reserved the right to read and sign the deposition;

THAT said deposition is a true record of the testimony given by the witness;

THAT, I am neither attorney nor counsel for, nor related to or employed by any of the parties to this action in which this deposition is taken and, further, that I am not a relative or employee of any attorney or counsel employed by the parties hereto, or financially interested in this action.

DATED THIS 13TH DAY OF SEPTEMBER, 1997.

JENNIFER S. SATI, RPR, CRR
Notary Public, Henn. County, Minn.
My Comm. Expires January 31, 2000

ERRATA SHEET

RE: Minnesota Tobacco Litigation

Scott Zeger, Ph.D., Volume I

I, SCOTT ZEGER, Ph.D., do hereby certify that I have read the foregoing transcript of the proceedings taken on September 10, 1997, and believe the same to be true and correct, except as follows:

PAGE LINE DESIRED CHANGE

Date ______________________ Notary ______________________

Date ______________________ Signature of Witness ______________________

PLEASE RETURN TO: Jennifer S. Sati, RPR, CRR
Ray J. Lerschen & Associates
620 Plymouth Building
12 South Sixth Street
Minneapolis, Minnesota 55402-1519